DOCUMENT RESUME

ED 195 928 CS 005 723

AUTHOR	Kirkwood, Kristian J.; Wolfe, Richard G.
TITLE	Matching Students and Reading Materials: A Cloze-Procedure Method for Assessing the Reading Ability of Students and the Readability of Textual Materials.
INSTITUTION	Ontario Inst. for Studies in Education, Toronto.
SPONS AGENCY	Ontario Dept. of Education, Toronto.
REPORT NO	ISBN-0-7743-5347-3
PUB DATE	80
NOTE	90p.
AVAILABLE FROM	Ontario Government Bookstore, 800 Bay St., Toronto, Ontario M7A 1L2, Canada ($3.00)
EDRS PRICE	MF01/PC04 Plus Postage.
DESCRIPTORS	Cloze Procedure; *Content Area Reading; Elementary Secondary Education; Foreign Countries; *Language Arts; *Readability; Reading Achievement; *Reading Materials; *Reading Research; Reading Tests; *Social Studies; Textbooks
IDENTIFIERS	Ontario

ABSTRACT

A study was conducted using standard open-ended cloze testing to establish a criterion for the utility of content reading passages for students in grades four, seven, and ten in the subject areas of language arts and social studies. Passages were selected from books, novels, and textbooks required in these courses. It was found that most of the reading passages were too difficult for most of the students and that, especially for students at the low end of the reading ability scale and in the lower grades, there were few appropriate materials at all. Social studies materials were found to be more difficult than language arts materials. Following a count of mean word size of the passages, a series of calibrating regression analyses were carried out to produce readability formulas that are on the same scales as the multiple-choice cloze test and the standard reading comprehension tests used in the schools. The formulas, when applied to a passage, estimate the test score a student would need to reach criterion on an open-ended cloze test or to understand the passage. (MKM)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED195928

U S DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

MATCHING STUDENTS AND READING MATERIALS

A Cloze-Procedure Method for Assessing the Reading Ability of Students and the Readability of Textual Materials

KRISTIAN J. KIRKWOOD, Principal Investigators
RICHARD G. WOLFE

FLORENCE MAYNES
JOHN MILLAR
KAREN SWORD
MARJORY SWORD

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
The Minister of Education, Ontario
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

This research report was funded under contract by the Ministry of Education, Ontario. It reflects the views of the authors and not necessarily those of the Ministry.

Hon. Bette Stephenson, M.D., Minister
Dr. H. K. Fisher, Deputy Minister

CS805723

© The Minister of Education, Ontario, 1980
Queen's Park
Toronto, Ontario

MINISTRY OF EDUCATION, CATALOGUING IN PUBLICATION

Kirkwood, Kristian J.
Matching students and reading materials: A Cloze-procedure method for assessing the reading ability of students and the readability of textual materials / Kristian J. Kirkwood and Richard G. Wolfe, principal investigators. --

This research project was funded under contract by the Ministry of Education, Ontario.

1. Reading - Ability testing. 2. Educational tests and measurements. I. Wolfe, Richard G. II. Ontario. Ministry of Education. III. Title.

ONO 1169 371.26

ISBN: 0-7743-5347-3

Additional copies may be ordered from:

Publications Sales
The Ontario Institute for Studies in Education
252 Bloor Street West
Toronto, Ontario
M5S 1V6

or from:

The Ontario Government Bookstore
880 Bay Street
Toronto, Ontario
M7A 1L2

1 2 3 4 5 XX 48 38 28 18 08

ABSTRACT

The investigators used standard open-ended cloze testing to establish a criterion for the utility of a reading passage for a student. A literacy survey ("literacy" is used in this study to mean matching the reader's skill with the reading material) was made to determine the overlap in the reading abilities of students and the language difficulty of reading materials in use in the schools. Also, a technology was created for matching materials and students.

The study focused on grades 4, 7, and 10, and on the subject areas of language arts and social studies. Four school boards participated in the study, with two providing data on language arts and the other two on social studies.

In language arts, all textbooks included in Circular 14 and in use in the two boards were selected for study, while in grade 10, a sampling was made of books and novels on the high school-approved reading list. Similarly, in social studies, a sampling was made of books listed in Circular 14 and in use in those boards. In each subject area and grade level, a total of thirty-nine passages were finally selected and prepared as open-ended cloze tests.

A major data collection was carried out in May-June, 1978. Over the four boards, 7201 students were sampled, with approximately equal numbers in each grade and board. All students wrote one form of a multiple-choice cloze test consisting of ten items borrowed from an earlier study [Leithwood et al., 1977]. Each student took a sample of three of the thirty-nine open-ended cloze tests for the grade and subject area. Overall, about one hundred responses were obtained for each open-ended cloze test. In the two language arts boards, students also took a standard reading comprehension test of the kind used in the boards' annual testing program.

The survey of literacy was established by tabulating the responses to the open-ended cloze tests.

It was found that most of the reading passages were too difficult for most of the students and that, especially for students at the low end of the reading ability scale and in the lower grades (4 and 7), there are few appropriate materials. Social studies materials were found to be more difficult than language arts materials.

Linguistic measures (e.g., mean word size) were made on the original passages. Following this, a series of calibrating regression analyses were carried out to produce readability formulas that are on the same scales as the multiple-choice cloze tests and the standard reading comprehension tests. The formulas, when applied to a passage, estimate the test score a student would need to reach criterion on open-ended cloze, that is, for the passage to have utility.

This study has implications for future procedures of matching students and textual materials.

A large proportion of the information that students are expected to acquire from about grade 4 in school is transmitted to them through written materials, especially textbooks, reference books, and supplementary reading books. In this connection, the role of the teacher and of educational administrators becomes one of facilitating reading - through selecting books, assigning reading passages, assisting students in reading - and of evaluating the student's understanding after reading. The success of the information transfer depends to a large extent on whether the students have sufficient linguistic ability to understand the language of the reading material or, conversely, on whether the materials are written at a level of linguistic complexity appropriate to the language ability of the students.

For purposes of this research, the matching of the skill of the reader and the difficulty or accessibility of materials is fundamentally what we mean by "literacy." In this sense, our research is concerned with the measurement and calculation of literacy in the schools.

It has long been part of educational practice to test student reading skill and to assign reading material to students according to their reading ability. The main contribution of this research is to put the two measurements-- the reading ability of the students and the difficulty level of reading materials--on a single, performance-based scale. The general plan of the project is given in Figure 1.

Literacy System

A complex system concerning literacy operates in the schools. First, a great amount of instructional time, especially at lower grade levels, is devoted to teaching reading. This time appears to be allocated in competition with the need for subject-area instruction. Second, through the publishing industry and the Ministry of Education, textbooks are written, published, approved, and made available. Teachers and administrators further evaluate books and select them for the instructional program. Third, teachers devote a large amount of instructional time in preparing students to read, evaluate, and consolidate their learning. Most of a student's time in school is allocated to reading.

One objective of this research was to study the functional capabilities of the literacy system in Ontario schools. Representative samples of students from four Ontario school boards and systematic samples of materials in use in those boards were drawn, and a survey was undertaken using the cloze procedure to compare the skill of the students and the difficulty of the materials.

The general conclusion is that there is a substantial mismatch: most of the books in use or available for use are too difficult for most of their potential readers. This will not surprise most classroom teachers [see Bormuth, 1977]. But the empirical demonstration of the mismatch and the identification of the grade levels and subject areas where the mismatch is more or less severe are important outcomes of the study for purposes of educational planning. With the knowledge gained from this study, it is now possible to see if additional instruction in reading skills would be beneficial. Further, we can now estimate how much instructional time must be taken away from subject -area instruction to prepare students to read the materials in those areas, and also determine whether more effort is needed in producing, selecting, and assigning books.

Another objective of this research was to develop a technology for improving the literacy arrangements in schools. General and specific surveys were carried out for direct examination of the readability of selected books for target groups of students. The cloze procedure proved to be effective in evaluating books for schools and systems. A by-product of the main survey was the development and cross-calibration of tests of student reading ability and formulas for estimating difficulty of materials. This is intended to facilitate the process of matching students and materials. Of course, teachers and administrators already spend a lot of their time and effort in selecting materials and matching them with students. However, the present research provides a systematic methodology for accomplishing the task with more precision.

Evaluating Literacy Performance

In the study of literacy, the basic unit of observation and analysis occurs when a student is presented with a passage to read. One immediate outcome is the student's attitude or willingness to read the passage. A second outcome is the amount of time which a student spends or needs to spend with a passage in order to read it. The main outcome, however, is the amount of information that the student gains from the passage. Attention has been restricted in this research to the literal comprehension of a passage, that is, to the student's ability to answer all questions that can be derived from the actual language in a passage. Information gain is defined as the number of those questions a student can answer after reading a passage in relation to the number the student could answer before.

The overall appropriateness or usefulness of a passage for a student represents a combination of information gain, time needed for reading, and attitude toward reading. Each of these components, and therefore the general utility of the passage, depends on the linguistic difficulty of the passage relative to the reading ability of the student. The exact connection has not been extensively researched in the current project; rather, we have

drawn on the work by J. R. Bormuth, who provides a weighting of the components and a set of criterion levels. These levels are expressed as percentage of correct responses on a standardized, open-ended cloze procedure.

To find the general utility of a passage for a student, an open-ended cloze test is constructed from the passage and the student responds by attempting to replace the exact missing words. The student's score on the cloze test defines the utility of the passage for the student. Bormuth holds that the minimum criterion for understanding is 35% and that a moderate level of response lies between 35 and 50%. He finds that when the response level falls below 35%, there is essentially no information gain. In addition, his recent work [personal communication, 1979] indicates that in this latter case the teacher has few possibilities of helping a student use the material. This new finding has major ramifications for future work in matching students and materials.

The definition of the utility of the student-passage connection, as conceptualized in terms of information gain, attitudes, and time, and as operationalized in terms of the cloze measurement, provides a tool for surveying the operation of the literacy system in the schools. In each of two subject areas - language arts and social studies - and at each of three grade levels - 4, 7, and 10 - thirty-nine passages were sampled from textbooks in use, cloze tests were constructed from the selected passages, and the tests were administered to representative samples of classes from the corresponding grades. The simple distribution of responses relative to the criterion levels of cloze aggregated over students and over books provided the basic assessment of the literacy system. This procedure yielded empirical evidence about what percentage of the books are accessible to what percentage of the students and about the range of materials available for poor readers and good readers in the schools.

A Technology of Literacy Evaluation

The survey undertaken also provided evidence about the readability of specific passages and books, and this kind of information can be useful in selecting books and, more importantly, in guiding the classroom use of such books. In fact, the open-ended cloze procedure can be used for system-wide or local evaluation and assignment of materials, and some experimentation with this approach was carried out in the present study. The steps for sampling passages, typing the cloze tests and keys, administering the tests with randomized item sampling, and scoring and tabulating the responses are straightforward and economical; a computer can prepare the tests and keys, and the students themselves can score their answers.

This procedure is quite feasible for relatively small numbers of books, passages, and students, but with larger numbers, it is important to avoid the large amount of testing time involved in

obtaining direct performance data on all passages of interest for all students.

Another aspect of the present research has been to provide general measurements on students and general measurements on passages. These can be linked to predict the performance of students on passages without the need to gather response data. The open-ended cloze criterion levels provide fixed points for relating the measurement scales. For example, a test score of 49 might be required for a student to reach criterion (35%) on a given passage. Conversely, a student might not be able to reach criterion except on passages with an average word length of less than, say, 5.5. In the process of carrying out the literacy survey, information on student reading skill, passage difficulty, and cloze performance was obtained to permit the cross-calibration of student tests with passage formulas [see Chapter 3].

Literal comprehension skill is one aspect of most standardized reading tests, and probably there is not much difference in correlation between the measurements provided by one test or another (or between these kinds of tests and general intelligence tests). [See Rankin, 1959; Hafner, 1963; Froese, 1971.] In this research both standardized reading tests and some experimental forms of a multiple-choice cloze test were used. The main problem with the conventional application of such tests is the arbitrary or normative kind of numerical scales which they produce. The goal in the present study was to produce a scale which has immediate practical use in classroom settings and to avoid reference to so-called "grade-equivalent" scores.

There exist a large number of "readability formulas" which are used by teachers, educators, and publishers. The readability information produced by one formula is highly correlated with that produced by another: differences between the word lists and the linguistic variables are minor [see Fry, 1962]. Deeper linguistic analysis could give more accurate and robust measures of readability, though for practical purposes the simple measurements are probably adequate. But, again, the problem with the conventional use of readability formulas is the arbitrariness of the numeric scales they yield. The numbers are not referenced to the skills of the students who are expected to read the material. It is very difficult to determine what student populations the "grade-equivalent" scores from a conventional readability formula refer to.

The key to a technology for literacy evaluation is a single scale for student reading skill measurement and for materials difficulty measurement. Like the traditional norm- referenced concept of grades or percentiles, the scale need have no particular numerical significance. But on the scale, a passage should have a score value equal to a student's score value if it is likely that the student given the passage under the standard open-ended cloze procedure would obtain a criterion level (35%). The student's number will be higher if the material is of moderate or low difficulty level for the student, or lower if the student, without

extensive help, finds the material inaccessible.

Applications to Educational Practice

In summary, this literacy evaluation research has yielded some immediately interpretable survey results and a set of calibrated measurement procedures which can be used in further practical applications.

It is a practical reality of the educational system in Ontario (and elsewhere) that the reading materials available are of variable difficulty in terms of language complexity. They are generally too difficult. This is never going to be completely remedied, and it is possible to argue that heterogeneity and various levels of difficulty are desirable. In any case, the survey and local cloze testing procedures, as well as the cross-calibrated student-material measurement procedures, can help teachers make better decisions about selecting and assigning textbooks and about determining the preparatory activities when materials of an ideal level of difficulty are not available.

References. The social and educational framework used in this study and described above is derived principally from the work of J. R. Bormuth of the University of Chicago. His notions about literacy and the literacy system operating in schools and in society is described most recently in Bormuth [1978]. The methodology used in the present research makes extensive use of the cloze procedure, which was first described by Taylor [1953]. Recent reviews of cloze testing are given by Jongsma [1971], Fram [1972], and Walter [1974]. Our particular approach is based on the cloze calculation studies reported by Bormuth [1971]. Similar literacy surveys have been carried out by Bormuth [1970] and Hanson [1974]. Further references to cloze testing and criterion measurement are given at the end of Chapter 1 of this report.

CONTENTS

1 Background/1
2 Project Methodology/9
3 Analysis and Results/15
4 Conclusions/23
Appendix/Examples of Cloze Tests/61
Bibliography/67

LIST OF TABLES

1. Size of Student Samples/27
2. Size of Materials Samples/27
3. Passage Assignment for Multiple-Choice Cloze Test Forms/28
4. Instrumentation Arrangements for the Main Survey/29
5. Distribution of Difficulty of Materials for Students of Different Levels of Reading Comprehension Skill, Language Arts, Grades 4, 7, and 10/30
6. Distribution of Difficulty of Materials for Students of Different Levels of Reading Comprehension Skill, Social Studies, Grades 4, 7, and 10/31
7. Overall Literacy Rates by Subject Area and Grade/32
8. Rescaling Values for the Multiple-Choice Cloze Test Forms/33
9. Summary of Passage-Level Correlations between Reading Comprehension Test Scores and Open-Ended Cloze Scores/34
10. Passage Difficulty as Calibrated into the Metric of the Reading Comprehension Test Scores, Language Arts, Grade 4/35
11. Passage Difficulty as Calibrated into the Metric of the Reading Comprehension Test Scores, Language Arts, Grade 7/36
12. Passage Difficulty as Calibrated into the Metric of the Reading Comprehension Test Scores, Language Arts, Grade 10/37
13. Passage Difficulty as Calibrated into the Metric of the Reading Comprehension Test Scores, Social Studies, Grade 4/38
14. Passage Difficulty as Calibrated into the Metric of the Reading Comprehension Test Scores, Social Studies, Grade 7/39
15. Passage Difficulty as Calibrated into the Metric of the Reading Comprehension Test Scores, Social Studies, Grade 10/40
16. Relative Values of the Regression Slopes for Average Word Length and Percentage on Non-Familiar Words/41
17. Summary of the Calibration Formula Regression Analysis/41
18. Calibration Formula Information/42
19. Valid Ranges for Calibration Formulas/43
20. Relative Values of the Regression Slopes of Word Difficulty and Sentence Length/44

LIST OF FIGURES

1. Overview of the Literacy Evaluation Project/45
2. Distribution of Difficulty of Materials for Students of Different Levels of Reading Comprehension Skill, Language Arts, Grade 4/46
3. Distribution of Difficulty of Materials for Students of Different Levels of Reading Comprehension Skill, Language Arts, Grade 7/47
4. Distribution of Difficulty of Materials for Students of Different Levels of Reading Comprehension Skill, Language Arts, Grade 10/48
5. Distribution of Difficulty of Materials for Students of Different Levels of Reading Comprehension Skill, Social Studies, Grade 4/49
6. Distribution of Difficulty of Materials for Students of Different Levels of Reading Comprehension Skill, Social Studies, Grade 7/50
7. Distribution of Difficulty of Materials for Students of Different Levels of Reading Comprehension Skill, Social Studies, Grade 10/51
8. Literacy Distribution of Passages by Students, Language Arts, Grade 4/53
9. Literacy Distribution of Passages by Students, Language Arts, Grade 7/54
10. Literacy Distribution of Passages by Students, Language Arts, Grade 10/55
11. Literacy Distribution of Passages by Students, Social Studies, Grade 4/56
12. Literacy Distribution of Passages by Students, Social Studies, Grade 7/57
13. Literacy Distribution of Passages by Students, Social Studies, Grade 10/58
14. How to Find the Gates-MacGinitie Test Score Estimated to Be Required for Literal Comprehension of a Passage in Grade 10, Language Arts/59

ACKNOWLEDGMENTS

A survey of this magnitude obviously requires the support and confidence of a host of people. We are indebted to the Ministry of Education, which funded the project. We would like to thank all of the principals, teachers, and students who took part in the reading survey. Without their positive co-operation, the study could not have been completed. Also, we would like to thank the superintendents and the co-ordinators of language arts of the four participating school boards for their support and encouragement.

Special recognition must be given to Dr. J. R. Bormuth, University of Chicago, for his contribution to the study.

K.J.K.
R.G.W.

1 BACKGROUND

The literacy survey procedures and the technology for literacy evaluation have their theoretical and substantive basis in current psycholinguistic theory. In particular, the cloze testing procedures which are employed are consistent with that theory, while conventional measurement procedures in the area of reading and readability have only indirect connection to what is considered the basis of reading. A general bibliography in the area of readability, psycholinguistic theory, and cloze testing is provided at the end of this report. The important concepts of the psycholinguistic model and its relationship to the use of the cloze procedure are discussed below. In addition, the conventional reading measurement procedures are discussed.

Psycholinguistic Model of Reading

Psycholinguistics, the combination of cognitive-developmental psychology and linguistics, provides an approach to understanding how students read and how they succeed or fail at extracting information from passages. The primary assumption of the psycholinguistic model is that the search for meaning is the real goal of reading. The process of reading is a cognitive process, in which the reader actively pursues the content or meaning in a passage. This implies that reading is not a simple auditory or visual transformation of print to speech, but rather it is complex, internal information processing, that is, a kind of thinking.

According to psycholinguists, the language represented in a printed text conveys meaning through three kinds of information: the graphophonic information of the marks on the page, the syntactic information, which is implicit in the order of the words of the text, and the semantic information, which derives from the combination of the words, the syntax, and the prior experience and knowledge of the reader. Because syntactic rules link the surface structure of letters on the page with the deep structure of the meaning of what is being read, the process of reading must place great stress on the relationships of words within sentences and of sentences within larger organizations [see Cooper & Petrosky, 1975].

Psycholinguistic research is based on the theory that fluent readers read for meaning, not just for identification of letters, words, or phrases. The fluent reader must use, in addition to graphophonic cues, syntactic cues from the text together with considerable prior knowledge of language and content to resolve the alternative meanings a sentence or paragraph can have. Beginning readers need more visual information from the printed page, and reading is slower for them because they do not have the stored

information and experience which would allow them rapidly to reduce the uncertainty of meaning with a minimum of visual information. Fluent readers need only a minimum of surface structure cues to arrive at the deep syntactic and semantic structure.

A fundamental assumption of research in psycholinguistics is that the brain is directing the eye rather than the eye directing the brain. That is, a reader establishes hypotheses on the basis of partial cues derived from the symbols on the page, makes hypotheses about potential meaning, predicts subsequent parts of the text, verifies hypotheses, and generally takes an active direction in working to an understanding of the meaning of the passage [see Cooper & Petrosky, 1976].

Implications for Readability Research

The psycholinguistic model for reading leads us to a refined basis for defining and measuring literal comprehension. Generally, evidence is required to substantiate that a reader can and does process the surface and deep levels of information in a text and arrives at a complete conceptualization of the syntactic and semantic structure of the text. Clearly, it will not suffice for a reader to recognize the words in a sentence or paragraph of text; the reader must capture the intricate syntactic connections between words, phrases, and sentences.

This definition of literal comprehension is operationalized in Bormuth's original cloze studies, which provide the basic justification for the use of the cloze criterion levels. Bormuth carried out a complete syntactic analysis and breakdown of the language in his passages. He identified each syntactic node and applied a linguistic transformation to the language structure around the node to turn the words into a question. The questions derived in this fashion can be answered by a reader of the passage if the reader has succeeded in processing the syntactic structure of the passage. The set of questions provides a complete examination of the comprehension of the passage. Bormuth's final criterion for information gain is the increase in the number of questions that the student can answer after reading the passage. (In fact, this had to be estimated by sampling questions and looking at parallel samples of students under "before" and "after" reading conditions.)

The psycholinguistic theory also suggests the kinds of measurements which are appropriate for assessing the comprehension skill of a reader and the reading comprehension difficulty of a text. For readers, it would be ideal to observe the speed and accuracy with which they process the information in a text and to know what array of language transformations they are capable of applying and what level of syntactic and conceptual complexity they are capable of absorbing. For texts, ideal measurement would involve examining the complexities in the complete syntactic structure and determining the amount and difficulty of information

processing necessary for deriving the deep structure from the surface content. These are the theoretically ideal measurement procedures, but, except for basic calibration studies, such as Bormuth's, practical applications require the use of some of the alternative approaches which are available.

One kind of alternative is to use simple measurements such as vocabulary knowledge of students and word length in texts. They are statistically correlated with the characteristics of reading ability and readability, and for practical purposes may be substituted for them [see Bormuth, 1966; Rankin, 1959]. This is basically why conventional measurement procedures are at all successful. Cloze testing comes closer to the ideal observation because it elicits the information processing activity and allows its progress to be observed.

Cloze Testing for Literal Comprehension

Generally speaking, cloze testing refers to any process in which students are asked to restore original text from text which has been altered or "mutilated" through word deletion. Performance on cloze tests is indexed by how much of the original text the student can recover [see Taylor, 1953].

There are various cloze testing procedures involving different deletion rates, use of systematic or random deletion, deletion of certain kinds of words, and so on. There are also various procedures for scoring the tests, including the acceptance of only exact replacements or the acceptance of synonyms for deleted words. The traditional open-ended cloze testing procedure was used in this study. In this procedure, every fifth word is deleted and only the exact replacements are accepted as correct responses. Bormuth [1977] has shown that different deletion procedures produce highly correlated scores. While the mean score levels differ according to the kind of deletion or the kind of acceptable replacements, the ranking of readers and the ranking of passages are relatively unaffected. For the purpose of this study the cloze procedure was chosen so that this work can be tied to the definitive calibrations of Bormuth.

It is important, however, to differentiate true cloze testing procedures from testing procedures which use deletions but provide students with a choice of possible substitutions. This form of cloze is called the multiple-choice cloze procedure. This procedure was utilized in the present study, but only as a means of providing normative-type measurement of the reading comprehension ability of students in the social studies part of the survey. The choice of the alternative distractors in the multiple-choice cloze procedure greatly influences the difficulty of each item and there is no simple way of predicting that difficulty except through collection of response information. Further, the nature of the response task no longer is a close simulation of the process of reading comprehension, and there exists no theoretical basis for linking

performance in the task to actual comprehension skill.

Traditional cloze procedures operate with mutilated text and require students to replace deleted words exactly. Cloze tests measure literal comprehension because the process of identifying the correct word simulates the process readers go through in reading the passage. Successful substitution for the deleted parts is possible because of the cues in the original language, but the students must use all the syntactic and semantic cues which are visible in the remaining graphic information. In other words, a reader must interact with the particular content and structure of the passage as far as it can be seen in order to consider possible substitutions. The cloze procedure measures the overall concept load and the overall syntactic and semantic processing that must be employed in reading and understanding the passage [see Taylor, 1953]. Because it depends on the apprehension of the deep syntactic and semantic structure of a passage, it is remarkably free of dependency on the surface differences in the styles of different writers [see Walter, 1974]. The standard open-ended cloze procedure provides a general measurement of the student's understanding of a passage. There seems to be disagreement among researchers about whether further examination of the cloze response patterns would provide diagnostic information about the reading problems and capabilities of individual students. Some recent work by Bormuth has demonstrated the correspondence between response to cloze tests and component skills that are involved in reading. He finds that students who respond at very low levels on a cloze test are demonstrating low-level reading skills, primarily the identification of words. When progressively higher cloze response rates are obtained, students are progressively demonstrating higher level skills: understanding of words, understanding sentences, understanding of anaphora and intersentence connection, and, ultimately, inference about intent and style.

In Bormuth's work, the technique of standard open-ended cloze testing has been shown to have a constant and consistent metric or measurement scale. Bormuth has mapped, over a broad range of both reading materials and student age and reading skills, the relationship between cloze test scores and the criterion outcomes of information gain, time spent in reading, and attitude toward reading. He found that a particular level of cloze response corresponds to a fixed amount of information gain, regardless of whether one deals with fluent readers of difficult passages or with poor readers of easy passages.

The criterion levels suggested by Bormuth have been employed in this study. Bormuth used the judgment of teachers about the relative importance of information gain, attitudes toward reading, and time spent in reading. He fixed a weighting of these three factors to define the general utility of the passage to the student. He then related this general utility to the cloze response level and determined the levels of cloze which optimize utility. He concluded that 35% of cloze response is the minimum for which a passage would be useful to students in terms of basic understanding. At anything less than this, there would be

effectively no information gain or else there would be an unwillingness on the part of the student to spend the time necessary to extract information. Bormuth found that 40% or 50% is probably ideal and that 50% and above meant that extracting information would be easier.

Conventional Measurement Procedures

The intent of this research has been to link the conventional measurement procedures for estimating materials difficulty and student comprehension through reference to the criterion performance as measured by standard open-ended cloze tests. This study does not claim any innovation in readability formula construction or in student comprehension test construction. The advance that this study provides is to put the student tests and materials evaluations on the same, performance-based scale. The following sections discuss some of the history and substantive issues in the conventional measurement areas.

Readability formulas. The measurement of the linguistic complexity of reading materials has a long tradition in reading research. Linguistic indicators, such as sentence length and word length, have been used extensively to classify reading materials. The differences between conventional approaches rest mainly in the particular definition of "familiar" words and in the weights assigned to different variables.

It is important to understand why readability formulas work. The variables of which they are composed do not, according to the psycholinguistic theory of reading, correspond to the complex information processing which a student uses in reading a passage. They refer much more to the surface structure of a passage than to the deep syntactic and semantic structure. Fortunately, in the natural population of conventional reading materials, there is a strong correlation between surface complexity and complexity in the deep structure [see Finn, 1978]. That is, when a writer has complex internal structure in terms of depth of syntax, amount of anaphoric reference, complexity of transformations,and so on, the writer usually also has longer sentences and bigger words. On the other hand, when a writer or editor reduces language complexity, the consequences include reduced sentence size and the use of more common and shorter words. From a psycholinguistic point of view, it is not the sentence length or the word length which is significant in increasing or decreasing the complexity; rather those characteristics are coincidentally related to real changes in the syntax and deep structure of the language.

There is a danger in conventional readability formulas. Their validity depends on their application to natural populations of reading materials, that is, materials that are not edited specifically to lower the apparent readability as measured by the surface variables in the formula. It is quite easy, for example, to transform a difficult text by cutting sentence length or reducing

average word size and the resulting edition would score lower on a readability formula. But the revision could provide no real change in the deep linguistic complexity of the language and therefore in the actual comprehensibility of the material. What in the original was complex intra-sentence syntax would become complex inter-sentence syntax in the modified material, and the information processing required of the reader would be the same. Good editors do more than alter surface structure.

It would be possible to develop more sophisticated linguistic variables that would tap the deep, linguistic causes of comprehension difficulty. They would, however, be more expensive to measure; for example, it might take a trained linguist to detect and count complex syntactic transformations. And the conventional finding of readability analysis is that for natural populations of materials, a few simple variables (especially the customary trio of word length, sentence length, and proportion of non-familiar words) seem to predict most of the significant variability in reading difficulty. This research has not gone beyond the standard, easily constructed linguistic measures.

Various studies on readability formula, such as the work by Dale and Chall, and Spache, use simple variables. The principal difficulty with their readability formulas lies in the way in which they ultimately use the variables to assign grade levels or numbers as estimates of materials difficulty. It should be noted that, in the absence of a performance criterion, there is no obvious or natural way to assign numbers. Given a weighting of the different variables, a rescaling could be applied to take the numbers into any desired range, from one to twelve or fifty to a hundred, or whatever. The initial work in readability analysis assigned numbers at "grade-equivalent" levels, and this was defined circularly by fixing the scales so that the average readability of the textbooks in a grade was equal numerically to the grade level. It is clearly appropriate only to the extent that the assignment of books to grade is rational and appropriate. Subsequent work in readability analysis has attempted to purify the variables, improve the weights, or confirm the assignment of numbers. The literature indicates that only with the recent work of Bormuth has the question of scaling readability formulas been put on a firm, empirical basis linked to the comprehension requirements of students [see Bormuth, 1977].

In summary, it is important to stress that the measures of readability derived in the present study will not differ much from conventional measures, at least in terms of correlation. But it can be expected that the scales for these numbers will correspond to the scales of the tests used in the research.

Comprehension skill tests. There are dozens of commercially available tests of reading comprehension skill, and countless numbers of locally developed and used testing procedures. An examination of the components of these kinds of tests shows that they are quite similar in form and content. Usually, a short paragraph is given to the students and they are asked three or four

questions which elicit various facts derived from the paragraph and perhaps the "main idea" of the paragraph or the "purpose" of the author. Conventional tests differ basically in the number of paragraphs they give the student and the specific nature of the questions asked. There are also some examples in standard tests of multiple-choice cloze testing or other somewhat more interactive procedures. All of these tests yield high intercorrelations and correlate quite highly with measures of general intelligence.

Because of the similarity, choice of a test should probably be based on technical reliability and cost. All available tests have severe problems in interpretation of scale, and it is this problem that this study is attempting to address.

It was decided to use the tests most widely employed in Ontario schools: the CTBS (Canadian Test of Basic Skills) in grades 4 and 7, the Gates-MacGinitie in grade 10.

In addition, several new forms of literal comprehension tests made up of multiple-choice cloze test items that had been prepared in an earlier project [Leithwood et al, 1977] were utilized. These items appeared to have considerable potential for practical use in Ontario schools, since they are constructed of familiar Ontario content, are cheap (cost of reproduction), and are easy and quick for students to take and score. Items for this particular form of cloze testing are constructed from materials such as local newspapers, magazines, and other familiar articles. Also, because there are a large number of items in the bank, it would appear easy to prepare parallel forms.

At the lower grades, the reading comprehension skills of the students can change rapidly, and in order for the teacher to continue to make accurate assignments of reading materials to students, they must continue to test the students' reading skills. For this purpose, experimentation was pursued with some simple, classroom-oriented testing procedures using the multiple-choice cloze passages. Because a general test of reading comprehension skill must include items covering a range of difficulty adequate for measuring the potential of all pupils, many of the items will be inappropriate for any individual student. The purpose of a sequential, tailored testing system is to ensure that, as far as possible, an individual student receives only items which are important for describing the student's reading skill. That is, one gives a student only items that are at just about the student's level, since only those items provide efficient information about the student's level. Experiments with systems for devising appropriate series of items to give a student have been carried out as well.

References. The interpretation of how cloze testing and readability measurement works is based on the psycholinguistic model for reading, reviewed recently by Goodman [1970] and Cooper and Petrosky [1976]. Smith [1973, 1975] and Taylor [1976] should also be consulted. The definitive calibration of the open-ended cloze procedure and the establishment of the criterion used here is from Bormuth [1971]. In our work we have followed Bormuth's standards exactly. Other deletion strategies have been proposed for purposes of classroom instruction using cloze [see Anderson, 1967, and Bortnick, 1975], and scoring procedures other than exact replacement have been studied [see Bormuth, 1965]. The correlative equivalence of the methods (scales shift but rankings are maintained) was shown by Bormuth [1965]. General reviews of cloze testing are given in Jongsma [1971] and Myers [1976]. Early work on readability formulas was completed by Dale and Chall [1948], Fry [1943,1948], Farr et al. [1951], Gunning [1952], Lorge [1959], McLaughlin [1969], and Fry [1968]. Good reviews are given in Klare [1965] and Guilliland [1972]. The connection between readability variables and cloze measures has been studied by Froese [1971] and exhaustively laid out by Bormuth [1978].

2 PROJECT METHODOLOGY

The principal analytic outcomes of this research are the "literacy survey " and the calibration of the literacy measurement procedures. Both are derived from a large-scale data collection effort made in May-June, 1978. This chapter provides the technical documentation of the design and the execution of the data collection.

Populations and Samples

The data collection was based on a joint sampling of students and reading materials. The collection of data could, in theory, have involved the books actually used in the individual classrooms. This would have led to an evaluation of the difficulties of reading experiences as they actually occur, given the presumably differential assignment of easier materials to poorer readers. However, the practical problems of making individual-level examination of reading experiences are overwhelming. Instead, in this survey, the sampling procedure was to draw independent samples of materials and samples of students and to estimate the potential for successful or difficult reading experiences.

Students. The results of the literacy survey part of the analysis were intended to illustrate the usefulness of the cloze procedure for purposes of studying the match of students and reading materials. The samples of students were designed to be statistically representative of the student population in the grade levels in the boards participating in the study. No claim can be made that these samples - students or materials - are representative of the province as a whole.

Participation of two boards of education was obtained for the language arts survey, and participation of two other boards was obtained for the social studies survey.

For the purposes of evaluating the readability of each passage, it was considered appropriate to have about one hundred students responding to the open-ended cloze test constructed from the passage, since that would give a minimum accuracy of about + or - 10% in the proportion reaching a criterion level of performance. Because thirty-nine passages were to be used in each subject area and grade level and each student was to respond to three passages, a target sample size of about thirteen hundred (= 39 x 100 / 3) per subject area and grade level was established. It was reasonable to have equal statistical precision in each of the school boards participating in the work, so the samples in each subject area surveyed were split equally between the two participating boards, producing a target sample size of about 650 cases per grade level

per board.

The sampling of students within a board was in most cases carried out in two stages: first, a sampling of schools, and then a sampling of students within schools. However, at the grade 10 level in the social studies part of the survey, all available schools were in the survey. The sampling of schools for language arts was done with probabilities proportional to the size of the grade, and in grades 4 and 7, additional stratification for grade size was made.

In all administrations, because of administrative convenience, the complete classrooms in grades 4 and 7 or complete English classes at grade 10 had to be used, so the sampling of students within school was done by random selection of classroom or classes as clusters. The final sample sizes for students and school are given in Table 1.

Reading Material. The open-ended cloze tests used in the main survey were constructed from passages drawn at random from the reading materials approved and in use at the grade levels and in the subject areas, though not necessarily in use in the individual classrooms. No attempt was made to sample according to frequency of use of materials, since that would have required a massive additional study of usage patterns. The population inferences from the results presented in the balance of this study are consequently phrased in terms of the reading materials available for use in these boards.

For each of the two subject areas and each of the three grade levels, the selection of reading materials passages was made by first selecting books and then drawing passages at random from the books. The final number of passages (thirty-nine in all cases) and numbers of books from which those passages are drawn are given in Table 2.

In the case of language arts, the grades 4 and 7 samples of books are, in fact, the populations, since there are relatively few approved language arts textbooks available at those grade levels and in use in the boards involved in this survey. For the grade 10 language arts survey, the books are a sample of thirty out of the several hundred books and novels that are on the approved English reading list for those boards.

In social studies, the Ministry of Education's Circular 14 lists many books as approved and potentially available for use. To eliminate books which are not in current use in the participating boards and to identify any books not listed in Circular 14 but in use in the two boards, a survey was carried out in which teachers and curriculum specialists were asked to indicate what books were in use. The books finally used in the main survey were drawn as a stratified random sample from the final list.

In both the language arts and social studies survey, after the books had been sampled, a passage was drawn at random from each one. From a random sample of the books, one or two additional passages were drawn so that the variability in difficulty within books could be evaluated along with the variability across books. In each subject and grade, enough additional passages were drawn to bring the total number of passages for the survey to thirty-nine.

Measuring Reading Ability

An earlier project by Leithwood et al. [1977] developed and pilot tested a pool of 257 multiple-choice cloze (MCC) items. One motivation for the present research was to provide further calibration of the MCC pool and to investigate possible procedures for practical application of the items in Ontario schools. Some effort was spent in two applications of the MCC pool which are not discussed further in this report: first, experimentation was carried out with the use of the 257 passages in a sequential, tailored testing procedure; second, computerized administration of the MCC items was programmed and tried out.

However, the main development of the MCC item pool was to create six forms of a reading comprehension test. Each form consisted of ten of the MCC items or a total of about ninety multiple-choice questions. Two parallel forms were developed for each of the grade levels 4, 7, and 10.

A reanalysis was carried out of the original response data which had been collected by Leithwood et al. in a multi-grade, multi-board survey two years previously. The overall difficulty level for each item was determined by utilizing the Rasch scaling technique. Further, the form and content of the passages were reviewed: items with technical errors or misprints were discarded as were items with particularly topical content, such as references to political events which had been current at the time of the original construction of the items.

The final list of usable items was classified from difficult to easy according to the Rasch scale values, and a stratified random selection was made. First, two medium-hard items were chosen to be used in all forms for purposes of equating the score scales. Then eight more items were drawn for each of the forms, 4A, 4B, 7A, 7B, 10A, and 10B. The two forms for each grade level were made randomly parallel, but the proportion of hard to easy items was increased for the higher grades. The final assignment of items to forms is given in Table 3 in terms of the item codes from Leithwood et al. [1977].

Passage Preparation

The survey design called for a total of 234 open-ended cloze tests

to be produced, because thirty-nine passages were to be used at each of three grade levels in each of two subject areas. Each cloze test is a full page of typewritten copy, and for purposes of accuracy and economy, it was necessary to computerize the production of the tests. This also facilitated later measurement of linguistic characteristics of the passages.

As described earlier, the passages were to be selected at random from specified or randomly selected books eligible for inclusion in the study. To select a passage from a book, a page was picked at random and then a line on the page was also picked at random. The passage commenced with the paragraph in which the random line was located. But if any essential information for interpreting the paragraph were contained in the previous paragraph, then the beginning of the previous paragraph was designated as the start of the selected passage. This continued, if necessary, to the beginning of a chapter. After the beginning of a passage was picked in this way, the next 250 words (minimum) were marked off. However, if a chapter or major section end intervened, the passage was kept short. Finally, the texts were typed into the OISE Decsystem-10 computer, and a careful proofreading was carried out.

Some further editing rules were applied when the materials were entered in the computer. Titles for the passages were invented by members of the project staff to give the reader a starting point for the passage. The titles were printed later at the tops of the cloze tests. Compound words were coded specially (using underscores instead of hyphens) when they were composed of bound morphemes; these would be deleted as a unit in the clozing process. Asterisks were used instead of periods in abbreviations, so they could be deleted in clozing. In order to avoid confusion between an apostrophe used in a contraction and a single quote, the single quotes were represented with another special character (#). Aside from these points, the texts were entered verbatim from the original material.

The construction of the open-ended cloze tests was carried out by a specially designed computer program (called CLOZE). The program skips over the title of a passage and then over the first nine words before making the first deletion. Then every fifth word is deleted until a total of forty-nine deletions are made or the text ends (see Appendix A). If there is more text after the forty-ninth deletion, the program copies the passage after the forty-ninth deletion until the first subsequent period indicating the end of a sentence. A separate program (called DCLOZE) operates the same way but produces a listing of deleted words, and this is used later in scoring responses to the cloze tests.

The texts were later combined into a single file for further linguistic analysis. This file contained the code designations of the passages, the titles (marked off with square brackets), and the texts as prepared for production of the cloze procedure.

Survey Procedures

The overall arrangements of assignment of instruments to students is indicated in Table 4. Each student received a random selection of one of the two multiple-choice cloze test forms that were available for the student's grade level. Each student also received three of the possible thirty-nine open-ended cloze test passages designated for use in the student's grade and board. The thirty-nine passages for a grade and subject area combination were distributed at random to the students with some controlled rotation for purposes of ensuring that each passage was more or less evenly distributed through the classroom and schools in the study.

All test administration in the language arts part of the survey was carried out by three members of the project staff. In all cases, the students wrote their tests in one sitting. The grade 10 students spent about twenty-five minutes doing the standardized reading test (Gates-MacGinitie), then twenty minutes with the multiple-choice cloze tests, and then forty-five minutes doing the three open-ended cloze passages. In grades 4 and 7, the timing was the same, except fifty-five minutes were given to the reading test (CTBS). The students were given breaks between the administration of the different parts of the survey.

In the social studies part of the survey, there was no standardized reading test other than the multiple-choice cloze form. The administration was carried out with assistants who were selected from substitute teacher lists in the counties. They attended a training session where they were familiarized with the open-ended cloze and multiple-choice cloze tests and the appropriate administrative techniques. Most of the grade 4 and grade 7 administrations were carried out in sequential sessions and regular classrooms. Where possible, the grade 10 students were tested in a single session in which they did both the multiple-choice cloze and open-ended cloze tests. In this part of the survey, there were no restrictive limits on the time students were given to spend in writing their tests; the administrators just waited until everyone was finished.

Only the standardized tests (Gates-MacGinitie or CTBS) in the language arts survey were able to be machine-scored. All other original scoring and collation of the data were done manually by project staff. The test booklets were numbered as they were received back. The open-ended cloze test passages, three per student, were separated into piles corresponding to the original thirty-nine passages per grade level and the number of correct responses was determined with reference to the computer-produced keys. The project staff carrying out this work was instructed to count only exact responses as correct, although they were told to be flexible in accepting spelling errors. Spelling errors were not accepted, however, when ambiguities were introduced (for example, "there" and "their").

The data processing resulted in two main data files for computer analysis, as reported in the following chapter. First, the main survey data were combined in a single file. This latter contained the student identification code, the designation of the school, board, classroom, the student's birthday and sex, the ten item scores for the multiple-choice cloze tests, and the three passage identification codes and numbers of correct responses for the open-ended cloze tests. A second file contained the actual texts of the reading passages used to create the open-ended cloze tests; this file was maintained for purposes of mechanical determination of linguistic variables. All files existed in two parallel forms, one for the language arts data and one for the social studies data.

3 ANALYSIS AND RESULTS

In this chapter we shall set out the results of the analysis of the main data collection of the project, together with explanations of how the analyses were carried out and interpretations of the results. The data analysis yields both the survey of literacy in the schools and the technology for literacy evaluation.

Literacy Survey Results

The basic survey of the literacy system as it operates in the schools of the participating boards, that is, of how well the students can read the books and how appropriate the books are for the students, comes from a tabulation of the open-ended cloze testing data. The passages from which the cloze tests were constructed are taken from books approved and in use in the grades and subject areas. The students responding to the cloze tests are a random sample of the students in the grade. The criterion levels of cloze response - 35% for separating hard from moderate and 50% for separating moderate from easy - supply the divisions for tabulating the responses. The following sections represent different ways of interpreting the joint tabulation of students and passages.

Literacy Distribution. A student in this study or anywhere in the schools has his or her own particular literacy level, as defined by the proportion of school materials the student would find hard, moderate, and easy. Some poor readers will find nearly all materials difficult, while the very best readers will find the majority of the materials easy. This is what is meant by the "distribution" of literacy. Because each student in the language arts survey took a general comprehension test, the open-ended cloze test responses could be broken down by student ability. In the social studies part of this survey, the multiple-choice cloze test was used to identify student ability. The results are presented in Tables 5 and 6 and in Figures 2 through 7.

[Technical note: separately for the language arts and social studies data, the test scores were ordered in each grade from low to high and eleven categories were chosen for tabulation. Then, all cloze responses were cross-tabulated with test score category, and the decile results presented in the tables and figures were derived by logistic interpolation. The multiple-choice cloze test forms, as rescaled, were used in this tabulation in social sciences, while the CTBS or Gates-MacGinitie scores were used in the language arts tabulation.]

The percentages in Tables 5 and 6 indicate how much of the material available in the different subject areas and grades is estimated to be hard, moderate, or easy for students in each decile

of reading comprehension skill. These are local deciles, that is, they refer to the bottom 10% through top 10% of the students in the participating boards. For example, from Table 6, the poorest 10% of the readers in grade 10 in these boards found 86% of the social studies materials to be hard, 12% to be moderate, and only 2% to be easy. On the other hand, the top 10% of the readers in grade 7 found 29% of the social studies materials in that grade hard, 44% of the materials moderate, and 27% easy.

The numbers in Tables 5 and 6 are graphed in Figures 2 to 7. The deciles, low to high, of reading skill level run across the base of the graph. Above each decile, the darker shading represents the proportion of hard materials, the lighter shading indicates the moderately difficult materials, and the white at the top indicates the easy materials.

One observation is that many books are difficult for many students. For the lower deciles, up to the fifth in all cases, the majority of the materials are difficult. Only for the highest deciles, representing the best readers, is there a substantial amount of moderate and easy materials. The materials have not changed; rather these students are better able to read them. Because the survey is based on the sample books approved and in use in these boards, but is not adjusted for frequency of use, these results indicate the possibility of finding useful material for students of different reading abilities. In all cases, the conclusion is seen to be that there are few materials for poor readers, or that poor readers are "literate" with only a small part of the available materials, particularly in grade 4 and in social studies.

Overall Literacy. The results by decile indicate the great need for careful assignment of reading materials to students of different reading abilities. The overall rates in Table 7 indicate what would happen if there were random selection of books. For each grade and subject area, the figure given in the table indicates how likely it is that a student drawn randomly from the grade would be able to perform at the criterion level (35%) of comprehension given a passage drawn randomly from the books in language arts or social studies from the grade. It is important to note that the overall rates go up with grade level. The students become better readers and the books become more difficult in the higher grades. The change in rates shows that the reading ability of the students and the readability of the materials converge at higher grades. But at all grades, and especially in social studies at grade 4, it is essential to take literacy into account when assigning books.

Analysis by Passage. The converse way of looking at the literacy distribution is through examination of the difficulty of particular passages. Figures 8 through 13 show the difficulty and ranking of the thirty-nine passages for each grade level and subject area. In each figure, the passages are ordered from easiest to hardest in terms of the estimated percentage of students who could meet the 35% cloze criterion. These percentages are given in numbers and in a bar graph, where shading means below criterion (hard) and white

means above criterion (moderate or easy).

[Technical note: for each passage, the response percentages for the one hundred or so students who filled out the open-ended cloze form were regressed on the students' general reading comprehension test score - that is, rescaled multiple-choice cloze in the social studies survey and CTBS or Gates-MacGinitie in the language arts survey - and adjusted to the total sample mean. Then the tabulation of students with adjusted scores above or below 35% was made.]

Again, there are a large number of students who performed below criterion. In language arts at grade 4, for example, only six of the passages were moderate or easy for more that half the students. In social studies at grade 4, there was only one passage which was moderate or easy for more than half of the students. The situation improves at the higher grade levels, so that by grade 10, more than half the passages were easy or moderate for more than half the students. However, there is still an apparent lack of easy reading material. At the other end of the scales, there are many passages which are moderately or easily accessible only for small percentages of the students.

The codes on the left side of Figures 8 through 13 identify the books and passages. The letters "a", "b", and "c" are used when several passages are drawn from the same book. For example, 707, 707a, 707b, and 707c would be four different passages selected at random from book 707. The numeric part of the code identifies the book. (It should be noted that the sample per book is very small, just one to three passages.) There is also substantial indication in this analysis of uneven language across a single book. For example, it is estimated that 36% of the grade 4 students can reach criterion performance on passage 425a, while none of the students can do so on passage 425b, yet both passages come from the same book. This suggests that either inadequate control was maintained over the language difficulty in the writing or editing of the book, or that there is intentional variation in readability, which is more likely in language arts than in social studies texts. The variability also means that complete readability classification of a book will have to be based on several samples.

Other Survey Results. In the sample design for the main survey, provision was made for separate analyses for the two participating boards in each subject area. Differences between boards were small but significant. It is especially interesting that some of the differences were reversed over grade levels, that is, literacy was higher in one board at grade 4 and in another at grade 10. Because these results are based on only four school boards, they cannot be generalized to cover all schools or boards.

Calibration Results

The open-ended cloze test responses that were tabulated for the

literacy survey results also provided the key to the cross-calibration of reading comprehension skill tests for students and readability formulas for materials. In this analysis, the single open-ended cloze criterion level of 35% was used. This means that the present study is relating student and materials measures at the point that Bormuth regards as the minimal level of performance; when student abilities fall short, or the language of texts is more complex, one would expect that substantial intervention by the teacher or extraordinary effort and motivation by the student would be required for reasonable success. It would be possible to reanalyse the data at a different criterion level of cloze, except that there are possibly not enough easy materials available in these subject areas and grades to permit a good calibration or suggest realistic matching at a higher level of performance.

The complete calibration involved several stages of analysis. First, for the social studies survey only, the two multiple-choice cloze forms for each grade were equated. Second, student performance on the open-ended cloze tests was translated into the scale of the reading ability tests. Third, the translated performance levels were regressed onto the linguistic characteristics of the passages. The results of the analyses are presented in Tables 16 through 20. The final product is a system for linking passage characteristics to test scores and is illustrated in Figure 14.

Standardizing the MCC Forms. There were two multiple-choice cloze (MCC) forms prepared for each grade level. Each consisted of ten items from the Leithwood et al. bank. While a complete scaling of the forms over grade levels, a rescaling of the whole MCC pool, and the development of the sequential testing procedure were all intended outcomes of this study, a serious problem was encountered. It was necessary to limit the reportable analysis of the MCC data to the within-grade equating of the pairs of forms and the calibrations of the forms with the readability formulas. In fact, the MCC items were found to be too easy. Even in the ninety or so multiple-choice questions which comprised one form, many students made just two or three mistakes. For the majority of items, most of the students made at most one or two errors. Part of the variation in scores could be attributed to careless answering, missing pages, or failing to finish the booklet. While the forms seem to measure reading comprehension, their measurement characteristics are very irregular because of the ceiling effects created by such easy items. The design of the forms included two common items which were in all six forms. One attempt at inter-grade equating involved Rasch scaling with an object of using the common items to link the forms. However, the relative difficulty of the common items reversed over grades. In another attempt, the common items were combined and regressed on the sums of the rest of the items, but the shapes of the regressions changed over form and over grade. The problem increases in the grades 7 and 10 forms. It should be noted that the forms constructed for the present study practically exhausted the hard items from the Leithwood et al. pool.

Also, it was found that, in the language arts part of the survey, sufficient time was not allotted to the administration of the MCC tests, so that a number of the students did not finish the tests. The data could have been analysed to equate for different administration times, but, given the other problems, that did not seem useful. Instead, it was decided to use just the standard reading tests (CTBS and Gates-MacGinitie) in the language arts survey and calibration.

However, for the social studies data, the MCC test data were all that were available, so an ad hoc equating had to be employed. The procedure was simply to equate the percentiles observed in the survey and then to convert the scales to a "normal" curve, which was arbitrarily set to run from 0 to 100. The rescaling values are given in Table 8. To use an MCC form, count the student's errors and look up the value for that number in the column corresponding to the form. For example, a student with five errors on form 7A would have a rescaled score of 69.

It is important to keep in mind that there is no connection between the scales given for forms in different grades, and that the two forms for a single grade have the same scale and should be interchangeable.

Calibrating Passages to Tests. Once the MCC passages had been rescaled, the calibration could proceed. It was decided to keep the CTBS and Gates-MacGinitie scores in their raw score (number correct) metric. Although the grades 4 and 7 students in the language arts part of the survey had written the same form of CTBS, the response levels were sufficiently separated for it to be unfeasible to try to pool that data. Instead, the calibration was done in six distinct analyses, one for each grade and subject area.

The basic procedure was to invert the regressions of open-ended cloze response to passages on the test score. A separate regression was calculated for each passage, based on the eighty to one hundred students who responded to the passage. In Table 9 the correlations obtained in these regressions are summarized. For each grade and subject area, there are thirty-nine passages and therefore thirty-nine regression analyses and correlations. The correlations are reasonably high and the regressions seem to be effective for all the 234 passages. From the equation CLOZE=A+B x TEST, a calculation is made to derive (35-A)/B, the test score at which the expected cloze is 35%. The numbers calculated in this way are measures of the difficulty of the passages, but they are scaled in the metric of the test.

The test-calibrated difficulty measures for the 234 passages are given in Tables 10 through 15, organized by subject area and grade. For example, in Table 10, it is estimated that a grade 4 student would need a score on the CTBS of at least 54 to attain 35% cloze on passage 403a or 408c. With a higher test score, we would predict better than criterion performance, but with a lower score, we would expect that the student could not reach criterion. The actual survey percentages of students performing above the 35%

cloze criterion, which for passages 403a and 408c are 24% and 30% respectively, are repeated in a parallel column.

[Technical note: the calibrated difficulties and the survey percentages do not correspond exactly, partly because of non-linearities in the cloze-test regressions, measurement error in cloze and the tests, and sampling error. The discrepancies are small.] In some cases, the regression estimate was suppressed because it fell off the test score scale or into a region of the scale where no passages were available to confirm the calibration.

It is important to note that the calibrated difficulty values are not comparable across grade or subject area, because the calibrating tests are different. The calibrated values can be used in evaluating the potential for using these passages with students for whom test score results are known. If a school or a teacher were to have these particular kinds of test scores available for the students, a matching of student scores with the values in these tables would provide key information about which materials are appropriate for which students, or about what degree of special preparation would be indicated when passages with high scores are given to students with low scores.

Readability Formula Derivation. Of course, in most cases, the practical school requirement will involve passages which are not on this list as well as a detailed examination of whole books. It will not often be practical to collect detailed student response data in such cases. The function of readability formulas, which are based on directly measured characteristics of passages, is to provide inexpensive readability information. Since the goal is to match students and books, the readability formulas have to refer in some way to student characteristics. In the case of this study, we would like them to refer to test score levels required to meet the 35% criterion of cloze performance. That is, the scales of the readability formula outputs should be exactly the same as the scales of the tests and the scales of the estimates given in Tables 10 through 15.

To achieve this correspondence, the final stage of the calibration processes involved regressing the calibrated passage difficulties on linguistic variables found in the language of the passages.

A recent review by Bormuth [1977] listed over one hundred reasonable linguistic variables which might be used to predict passage difficulty. Some are very simple and mechanical, such as average word size; others involve counting deep linguistic transformations and would require a skilled linguist for measurement. For practical purposes, a few simple variables have been found to account for most of the variation in the difficulty of ordinary textual materials. In this research, we also wanted to restrict our attention to measures that could be easily programmed for computer analysis. The analysis is limited to three variables: average sentence length, average word length, and percentage of difficult words. The list of non-difficult (familiar) words was

taken from the Dale-Chall readability procedure and word list, except no special provision was made for proper names. A computer program (called LINGV) was written to read the data file with the passage texts, as used in the cloze operation, and to count words, sentences, letters, and hard words. The processing is simple except that alternate forms, such as plurals, of common words have to be recognized as common. The processing time is about ten seconds per two-hundred-and-fifty-word passage on the OISE Decsystem-10 computer.

For each grade and subject area, the data now consisted of recordings on calibrated passage difficulty, in the test score metric, and average sentence length, average word size, and percentage of difficult words. The object was to regress passage difficulty on the linguistic variables. However, average word size and percentage of difficult words were found to have a very high correlation, about .7 or .8. This created a co-linearity in the regressions and caused a great instability in estimating the effect of either variable. The solution adopted was to combine the variables into a single index, called "word difficulty." The relative weighting in the index was decided after examining the regression coefficients in separate regressions with sentence length and word size and then with sentence length and difficult words. Table 16 represents the comparison of regression weights, standardizing by the slope of sentence length. The ratio is reasonably close to twenty in all cases, so that the weighting in the index was fixed at twenty (for word length) to one (for difficult words).

Another problem was that a few passages had very long sentences, and preliminary regression analysis showed that the presence of very high average sentence length values was greatly distorting the regressions. This was solved by redefining sentence length to be words per sentence or forty, whichever is less. This statistical procedure is known as "Windsorizing."

The final readability formulas were then calculated by doing a regression analysis for each grade and subject area. The dependent variable was calibrated passage difficulty, in the metric of the test scores, and the independent variables were average sentence length and the word difficulty index. The multiple correlation coefficients for these regressions are given in Table 17. They are high, especially for the social studies data. This means that the formulas produce estimates of reading performance which closely match the observed performance of students. Tests for outliers, non-linearities, and interactions revealed no further problems.

The calibrated readability formulas are given in Table 18. For each grade and subject area, the formula is defined by a weight for average sentence length, a weight for the word difficulty index, and a constant. Reference is also made to the test score scale to which the formula is calibrated.

An example of the application of the formula system is given in Figure 17. Here we suppose that a passage is drawn which has 276 words, 12 sentences, 1687 letters, and 81 hard words. From these counts, which can be made by the computer program LINGV, the variables of average sentence length, average word length, and proportion of hard words can be calculated by taking the appropriate ratios of the raw counts; for this passage, we get 23.00 words per sentence, 6.11 letters per word, and 29.35% hard words. (If the word/sentence ratio had been more than 40, we would have used 40.) The index of word difficulty is calculated by multiplying the word size by 20 and adding the proportion of hard words, and this gives 151.59. Finally, the weights from Table 18 are applied and the readability value is found to be 49.71. This means that a grade ten student would require a raw score of 49.71 on the Gates-MacGinitie to score the minimum criterion (35%) on a cloze test. Obviously, only the best students can be expected independently to use the passage.

Sometimes the formula will yield values which are higher or lower than the test score allows. This means that the passage is very difficult or very easy. While that indication is valid, the numeric value should not be interpreted. It is also possible that a value may be obtained which is in the valid score range for the test but which is beyond the range of passages encountered in this calibration analysis. Again, the indication of great difficulty or easiness may be believed, but the actual numbers should not be interpreted too strictly. Table 19 gives a range of values that should be accepted for each grade and subject area.

In the design of this research, it was not anticipated that the result would be six different readability formulas. This is partly due to the problems associated with arriving at a single scale for MCC items. A careful examination of the formulas as derived for the different grades and subject areas reveals that a single formula would in fact be wrong! This is a rather surprising result and explains possibly why conventional formulas have produced problems in reading research. The significant observation is that the relative regression weights of sentence length and word difficulty vary between grades and between subject areas. The ratio of the coefficients is given in Table 20. Generally, the relative importance of word difficulty compared to sentence length is greater in language arts materials than in social studies materials. The relative importance of word length compared to sentence length decreases as we move from grades 4 to 7 to 10 in either language arts or social studies. A formula that works at one grade or in one subject area may produce quite distorted estimates if applied to another grade or area.

4 CONCLUSIONS

This project had two main objectives. The first was to survey the existing match between student reading abilities and the difficulty of reading materials in selected grades and subject areas. The results of this survey have implications for allocating resources in reading instruction, for producing and selecting reading materials, and for understanding the impact of reading difficulty on instruction in other subject areas. The second objective has been to develop a technology for literacy evaluation that would facilitate system-wide and classroom matching of materials and students. This work is important for improving the utilization of existing reading materials. We feel that the research has also clarified some needs for further study in the field of literacy.

We hope that our readers will not try to extrapolate our findings to reading in the grades below grade 4, or to see these procedures and results as significant to the teaching of reading or the diagnosis of reading problems. We have been concerned instead with reading as it forms a part of instruction in the higher grades (4, 7, 10) in school and with materials that are intended primarily for instruction. The Ministry of Education has a program of distributing sample copies of new texts to schools. Our work describes methods for estimating the difficulty of materials that might be selected by educators.

While the four boards in this survey are not representative of school boards in the province, the results of the survey are striking. Especially at lower grades, and particularly in social studies, few of the reading materials available in the boards for use were accessible to the poorer readers. And yet the pattern of much of the instruction in schools requires students to gain much of their information from reading. Clearly, in some cases the information is not readily accessible, and in many cases, large amounts of teacher time and student time must be dedicated to special work for using material that would otherwise be too hard. We are not talking here about a small fraction of the students who might have pathological reading problems; rather we are dealing with the bottom half, or even, in some cases, with more than half of the students. We found that the language used in the textbooks surveyed was inaccessible to a substantial proportion of the students.

While the results are striking, they will not be surprising or perhaps even alarming to many teachers. One teacher who assisted in part of this research noticed that a recently published textbook that this analysis found to be difficult was, in fact, the book she had just stopped using as a textbook because her students could not read it. She was continuing to use it as a reference book and was pleased with it for that purpose. Teachers spend a lot of time gathering data and making judgments about the appropriateness of books for their students.

On the other hand, the overall results of the literacy survey suggest that between the teaching of reading, the production and the selection of books, and the use of reading, there is a problem of matching, especially in other subject areas. The results of the present study pose several questions which might be considered in future research. For example: Why are there so many hard materials and so few easy ones? Why should social studies instruction have reading difficulty as a substantial obstacle? What procedures should be used for monitoring the levels of readability and selecting of textbooks? Are teachers aware of readability in all subject areas? The information provided by this survey could help promote discussion and examination of these questions.

A reasonable direction for further research would be to extend the literacy survey to other grades and subject areas, to increase the coverage of the approved book lists, and to consider regional or provincial student samples. The production of the cloze tests is now automated. A sample of about one hundred students per test is adequate. In some other work, we have found that students can score their own (or their seatmates') work. For these reasons, an extended survey could be carried out at fairly low cost. Since the variability can be high within books, one needs several passages per book; also some combined strategy of data collection on cloze responses and of the application of a formula for linguistic complexity might be most economical.

Changes in instruction and changes in materials production are not going to produce great changes in literacy, at least not immediately. This research offers another kind of solution: namely, certain tests and formulas which can assist teachers and administrators in making better decisions about the assignment and use of written materials. One possibility is to use cloze testing directly in classroom or school situations. This is feasible but becomes expensive and time-consuming when dealing with many students or many passages. A practical improvement is to use cloze testing to calibrate passages against test metrics and then use test scores on students to determine potential match. A third possibility is an approximation, but it avoids all collection of response data, except for calibration studies like the present one: formulas are defined that operate on linguistic measures made on passages and produce estimates of the test score requirements.

Our results and experiences show the feasibility of these kinds of matching. The estimation procedures seem quite accurate. Of course, these immediate results are limited to a few tests and a few subject areas and grades. But the pattern of further research for extending the technology is clear. Representative student samples are not necessary, as long as a reasonable range of student ability is obtained. The selection of passages for clozing should be determined purely by the anticipated application of the evaluation system: one should use adequate samples from books which are likely to be available and used. The student reading comprehension tests should be exactly the ones in use in the schools. In that case, the survey and analysis procedures would parallel the present study, as described in this report. Computer

programs can be made available for producing the cloze tests and keys, and carrying out the analysis. The results would include tables of passage difficulty calibrated to the test scores. Further research might follow the pattern of this data collection and use actual random samples of students and passages. If that were done, the calibration study would also serve as a literacy survey, that is, a study of the match between students and textual materials.

In this study, the prediction of passage difficulty from linguistic variables proved to be accurate, and the approach to classifying materials seems quite reasonable. In future studies, it can be recommended that simple linguistic measures be used to extend the sizes of the passage samples. This would improve the accuracy of book evaluations without increasing the response samples. In addition, the readability formulas constructed in a calibration study can be used to classify new materials, at least from the same grades and subject areas.

However, the lack of homogeneity of the readability formula regressions over subject area and grade suggests that formula construction has some major problems. It is logical to suppose that the inconsistencies arise because the variables are so much a function of the surface structure of the language. Psycholinguistic theory would suggest that by measuring deeper characteristics of language which are more dependent on the information processing that must take place during reading, we might obtain formulas which operate over a wider range of content and level. Further research is needed to examine new variables and to test their performance. We must certainly conclude from these data that traditional readability formulas cannot be producing accurate results over the wide range of materials to which they are conventionally applied.

In this study only limited success has been achieved in preparing new instrumentation for measurement of individual skill in reading comprehension. It did not prove possible to link the MCC passages over grades, and, especially at higher grades, the tests did not discriminate because they were too easy.

Testing of reading comprehension skill is, in theory, quite simple and robust; all valid and reliable measures will be highly intercorrelated and highly correlated with general intelligence. The design of new instrumentation should aim at accuracy, with reasonable cost, and should result in useful calibration of materials. We feel that the development of new tests and testing strategies is a cost-effective direction for new research.

TABLES

Table 1/Sizes of Student Samples

	Grade 4		Grade 7		Grade 10	
	Students	Schools	Students	Schools	Students	Schools
Language Arts Survey						
Board 1	515	14	626	10	694	10
Board 2	661	16	595	10	558	8
Total	1176	30	1221	20	1252	18
Social Studies Survey						
Board 3	605	16	611	12	552	7
Board 4	645	24	615	11	524	8
Total	1250	40	1226	23	1076	15

Table 2/Sizes of Materials Samples

Language Arts Survey

	Grade 4	Grade 7	Grade 10
Books	25	21	30
Passages	39	39	39

Social Studies Survey

	Grade 4	Grade 7	Grade 10
Books	26	26	26
Passages	39	39	39

Table 3/Passage Assignment for Multiple-Choice Cloze Test Forms

	FORM					
	4A	4B	7A	7B	10A	10B
Position						
1	034011	035411	058311	036611	045011	041411
2	069311	069311	069311	069311	069311	069311
3	033311	042811	041811	056012	059021	070012
4	053031	055412	072732	053431	082741	075631
5	056431	032011	062031	041412	074811	094421
6	039011	058313	067731	058031	080721	076541
7	072811	072811	072811	072811	072811	072811
8	066231	056013	081021	081231	074931	078671
9	063411	055411	101831	071312	077012	077251
10	063611	075671	088951	080751	083751	078751

Table 4/Instrumentation Arrangements for the Main Survey

	In each grade and subject area, 39 passages of 250 words each were sampled from books in use, and open-ended cloze tests were prepared from them. Each student responded to 3 of the 39 tests.	For each grade, 2 forms of 10-passage multiple-choice cloze tests were constructed. Each student responded to 1 of the 2 tests.	Some of the students also took part of a standard reading test.
Language Arts Survey			
Grade 4	3 of the 39 grade 4 Language Arts passages	Form 4A or 4B	CTBS Form 4M (Reading Comprehension)
Grade 7	3 of the 39 grade 7 Language Arts passages	Form 7A or 7B	CTBS Form 4M (Reading Comprehension)
Grade 10	3 of the 39 grade 10 Language Arts passages	Form 10A or 10B	Gates-MacGinitie Form F1 (Comprehension)
Social Studies Survey			
Grade 4	3 of the 39 grade 4 Social Studies passages	Form 4A or 4B	-
Grade 7	3 of the 39 grade 7 Social Studies passages	Form 7A or 7B	-
Grade 10	3 of the 39 grade 10 Social Studies passages	Form 10A or 10B	-

Note: In addition, all students responded to a general information sheet indicating age, sex, and school and classroom assignment.

Table 5/Distribution of Difficulty of Materials for Students of Different Levels of Reading Comprehension Skill Language Arts, Grades 4, 7, and 10

Local decile on reading test	GRADE 4			GRADE 7			GRADE 10		
	Hard	Moderate	Easy	Hard	Moderate	Easy	Hard	Moderate	Easy
1 (low)	93	6	1	88	10	1	73	23	4
2	93	6	1	81	15	4	68	26	7
3	91	8	2	76	19	4	62	31	7
4	82	15	3	75	19	6	54	36	10
5	80	17	4	72	20	8	47	39	14
6	72	25	3	65	25	10	44	40	16
7	65	29	5	55	29	16	39	39	22
8	60	30	10	50	31	19	32	43	25
9	52	31	17	41	34	25	24	42	34
10 (high)	38	38	24	24	42	34	12	36	52

Table 6/Distribution of Difficulty of Materials for Students of Different Levels of Reading Comprehension Skill, Social Studies, Grades 4, 7, and 10

Local decile on reading test	GRADE 4			GRADE 7			GRADE 10		
	Hard	Moderate	Easy	Hard	Moderate	Easy	Hard	Moderate	Easy
1 (low)	99	1	0	93	7	0	86	12	2
2	97	3	0	89	10	1	77	19	4
3	97	3	0	84	14	2	70	26	3
4	94	5	1	79	18	2	57	35	8
5	90	8	2	74	23	3	49	42	9
6	87	11	2	61	34	4	45	41	14
7	83	15	2	55	40	6	36	43	21
8	76	19	4	50	40	11	24	52	25
9	68	25	6	36	45	19	24	41	35
10 (high)	56	29	15	29	44	27	10	40	49

Table 7/Overall Literacy Rates by Subject Area and Grade

	Overall Literacy Rate
Language Arts	
Grade 4	28%
Grade 7	38%
Grade 10	57%
Social Studies	
Grade 4	15%
Grade 7	37%
Grade 10	53%

Note: Each rate represents the probability of meeting the 35% criterion level of cloze for a student chosen at random from the grade given a passage drawn at random from the materials in the subject area for the grade.

Table 8/ Rescaling Values for the Multiple-Choice Cloze Test Forms

Number of Errors	Form 4A	Form 4B	Form 7A	Form 7B	Form 10A	Form 10B
0	99	99	99	99	94	99
1	99	99	99	88	80	99
2	97	99	90	81	72	85
3	90	99	81	75	65	75
4	87	98	74	70	60	69
5	82	95	69	66	56	65
6	78	91	66	63	52	62
7	75	86	63	61	49	59
8	72	82	61	59	46	57
9	69	78	59	57	44	54
10	67	75	57	55	42	52
11	64	72	55	53	40	50
12	62	70	53	51	38	48
13	60	68	51	50	37	47
14	59	66	50	48	35	45
15	58	65	48	47	34	44
16	56	63	47	46	33	42
17	55	62	45	45	31	40
18	54	61	44	44	30	39
19	53	60	43	43	29	38
20	52	59	42	42	28	37
21	51	57	41	40	27	36
22	50	56	40	39	26	34
23	49	55	39	38	24	33
24	48	53	37	37	23	31
25	47	52	36	36	21	30
26	46	51	35	34	21	29
27	45	50	34	33	20	28
28	44	49	33	32	20	27
29	44	49	33	31	19	26
30	43	48	32	30	18	26
31	42	46	31	29	18	25
32	42	46	30	28	17	23
33	41	45	29	27	15	21
34	40	44	28	26	14	20
35	[illegible]	43	27	25	13	20
36	39	42	26	24	13	19
37	39	41	25	23	13	18
38	38	41	25	22	13	17
39	37	40	23	21	13	16
40	37	39	22	19	13	16
41	36	38	21	19	12	15
42	35	38	20	19	11	15
43	34	37	18	19	11	14
44	34	36	16	18	11	12

Note: at each grade level, the A and B forms are on the same scale, but there is no substantive connection in the scales between grades.

Table 9/Summary of Passage-level Correlations between Reading Comprehension Test Scores and Open-ended Cloze Scores

	Average Correlation	Range of Correlations
Language Arts		
Grade 4	.50	.27 - .65
Grade 7	.58	.41 - .70
Grade 10	.55	.37 - .70
Social Studies		
Grade 4	.61	.48 - .73
Grade 7	.65	.47 - .76
Grade 10	.61	.39 - .72

Note: For each subject area and grade, there were 39 correlations, one for each passage. Each correlation was based on a sample of about 80 to 100 students.

Table 10/Passage Difficulty as Calibrated into the Metric of the Reading Comprehension Test Scores

Language Arts, Grade 4

Passage Number	Estimated Test Score to Meet Criterion	Percentage in Survey Actually Above Criterion
403b	**	4
411	**	9
413a	**	11
417b	**	7
418a	**	14
419	**	18
420	**	4
421	**	5
424	**	0
425b	**	0
418b	66	18
423	66	13
415a	63	15
416a	63	29
405b	62	14
416b	59	16
422	56	21
410	55	24
412	55	30
403a	54	25
408c	54	23
417a	53	32
406b	52	24
401a	51	37
425a	50	36
413b	49	40
404	49	35
405a	49	36
407	48	33
408b	44	43
409a	43	49
408a	43	47
402	43	53
415b	43	48
406c	42	55
414	42	52
401b	40	57
406a	39	57
409b	34	60

Note: "**" appears when the required test score is off the range of the test or when no passages were found to establish the calibration. The test is CTBS Form 4M, Reading Comprehension.

Table 11/Passage Difficulty as Calibrated into the Metric of the Reading Comprehension Test Scores

Language Arts, Grade 7

Passage Number	Estimated Test Score to Meet Criterion	Percentage in Survey Actually Above Criterion
706	**	7
706a	**	7
707	**	5
718	**	7
702	75	11
705a	74	9
708	70	15
712b	70	13
710b	69	14
717a	69	20
703	64	23
716	64	28
717	61	19
703a	61	22
711	60	17
711b	59	21
720	57	22
721a	57	24
711a	55	31
704	54	38
701	53	32
712a	52	42
716a	52	39
719	50	37
714	49	38
709	47	50
718a	47	53
712	46	46
710a	46	62
707b	42	59
713	40	64
710	39	67
705	38	72
704a	38	68
721	33	70
715	30	69
715a	30	74
707a	29	80
714a	13	89

Note: "**" appears when the required test score is off the range of the test or when no passages were found to establish the calibration. The test is CTBS Form 4M, Reading Comprehension.

Table 12/Passage Difficulty as Calibrated into the Metric of the Reading Comprehension Test Scores

Language Arts, Grade 10

Passage Number	Estimated Test Score to Meet Criterion	Percentage in Survey Actually Above Criterion
1024	**	2
1011	49	9
1028b	44	23
1003	43	14
1012	42	25
1007	38	34
1026b	37	23
1008	36	34
1021	35	33
1028a	33	42
1014	32	36
1006	32	38
1018	31	51
1022	31	47
1004	30	41
1027	29	53
1025c	26	55
1017	25	52
1020	25	61
1029	25	57
1023c	24	62
1005	23	74
1030a	23	71
1016	22	64
1019	22	73
1025a	21	64
1025b	21	67
1030b	21	71
1026c	20	71
1009	17	76
1023b	16	79
1013	16	79
1026a	14	82
1002	13	83
1010	9	93
1015a	9	84
1001	6	92
1015b	**	96
1023a	**	99

Note: "**" appears when the required test score is off the range of the test or when no passages were found to establish the calibration. The test is Gates-MacGinitie Form 1F, Comprehension.

Table 13/Passage Difficulty as Calibrated into the Metric of the Reading Comprehension Test Scores

Social Studies, Grade 4

Passage Number	Estimated Test Score to Meet Criterion	Percentage in Survey Actually Above Criterion
476b	***	1
475b	***	1
475a	***	1
468	***	5
469a	***	2
476a	***	1
461b	***	2
454a	***	1
453b	***	1
453a	***	0
451	***	1
463	97	2
464b	97	4
465b	97	6
452b	97	5
452a	97	3
460	95	4
466	87	11
459	85	13
458b	85	14
456	83	14
457	82	13
464a	82	17
454b	79	23
462b	77	13
462a	76	22
461a	73	18
458a	73	21
472	73	23
455	68	30
469b	68	27
473a	67	27
470a	67	28
467	66	27
470b	66	23
473b	61	36
474	58	44
471	55	45
465a	48	59

Note: "***" appears when the required test score is off the range of the test or when no passages were found to establish the calibration. The test is MCC form 4A or 4B, rescaled.

Table 14/Passage Difficulty as Calibrated into the Metric of the Reading Comprehension Test Scores

Social Studies, Grade 7

Passage Number	Estimated Test Score to Meet Criterion	Percentage in Survey Actually Above Criterion
770b	**	0
771	**	2
764	**	3
766	**	4
770a	89	6
757b	84	9
776a	83	9
762a	78	14
751	78	13
768a	77	14
753	75	11
756	70	23
772b	70	17
754	69	23
760b	64	40
767	64	33
768b	63	32
759a	61	27
776b	60	38
757a	59	43
758b	59	43
769a	57	45
763	57	49
758a	55	37
755	54	47
762b	53	49
759b	52	42
760a	51	48
775a	50	52
769b	50	46
774a	50	53
761b	48	62
765	47	68
761a	47	61
775b	41	66
752	40	67
774b	38	70
772a	28	80
773	21	88

Note: "**" appears when the required test score is off the range of the test or when no passages were found to establish the calibration. The test is MCC form 7A or 7B, rescaled.

Table 15/Passage Difficulty as Calibrated into the Metric of the Reading Comprehension Test Scores

Social Studies, Grade 10

Passage Number	Estimated Test Score to Meet Criterion	Percentage in Survey Actually Above Criterion
1076a	**	2
1067	81	17
1057b	79	18
1076b	75	13
1071a	74	15
1071b	69	28
1063	64	27
1064	61	37
1070	61	34
1054a	59	38
1057a	57	37
1056	57	42
1069	55	51
1058b	54	43
1053b	52	46
1061b	52	53
1051a	50	60
1051b	50	47
1053a	49	52
1060	47	56
1052a	47	60
1072	47	56
1055a	45	63
1073	45	64
1054b	45	69
1075	45	68
1066	43	60
1068b	40	69
1062	39	73
1061a	39	75
1065a	37	78
1068a	36	66
1065b	36	80
1074a	35	79
1058a	34	76
1055b	34	78
1074b	32	79
1052b	28	83
1059	16	90

Note: "**" appears when the required test score is off the range of the test or when no passages were found to establish the calibration. The test is MCC form 10A or 10B, rescaled.

Table 16/Relative Values of the Regression Slopes for Average Word Length and Percentage of Non-Familiar Words

	Language Arts	Social Studies
Grade 4	26	22
Grade 7	14	19
Grade 10	21	11

Note: The figures represent the relative regression slopes of Average Word Length and Percentage of Non-Familiar Words in predicting the difficulty of a passage (measured by test scores needed to reach criterion cloze). Each comparison is based on two regression analyses over the 39 passages for the subject area and grade. Sentence Length was also a variable and its slope was used to standardize the slopes of these variables before the comparison. The conclusion is that the impact of one letter per word is about twenty times the impact of one percent of difficult words.

Table 17/Summary of the Calibration Formula Regression Analyses

	Language Arts	Social Studies
Grade 4	.59	.78
Grade 7	.53	.64
Grade 10	.56	.75

Note: Each figure is the multiple correlation in a regression over 39 passages in the particular grade and subject area. The dependent variable was estimated test score necessary for reaching the 35% cloze criterion, and the independent variables were Sentence Length and Word Difficulty Index.

Table 18/Calibration Formula Information

	Weight for Sentence Length	Weight for Word Difficulty Index	Constant Term	Calibrating Test Instrument
Language Arts				
Grade 4	.46	1.07	-49.70	CTBS Form 4M (Reading Comprehension)
Grade 7	.83	.81	-41.32	CTBS Form 4M (Reading Comprehension)
Grade 10	.87	.38	-27.91	Gates-MacGinitie Form 1F (Comprehension)
Social Studies				
Grade 4	1.16	0.95	-34.60	MCC Form 4A or 4B, Rescaled
Grade 7	2.22	0.45	-29.89	MCC Form 7A or 7B, Rescaled
Grade 10	1.21	0.24	-19.78	MCC Form 10A or 10B, Rescaled

Note: These weights and constant term define a formula for Sentence Length and Word Difficulty Index which yields the test score estimated as necessary to reach the 35% cloze criterion. Word Difficulty is Percentage of Non-Familiar Words plus twenty times Average Word Length.

Table 19/Valid Ranges for Calibration Formulas

	Lower Bound	Upper Bound
Language Arts		
Grade 4	30	65
Grade 7	20	75
Grade 10	5	50
Social Studies		
Grade 4	45	99
Grade 7	20	90
Grade 10	20	90

Note: These are the lowest and highest values that should be used out of the calibration formulas. Values outside these ranges are either off the test score scale or cannot be substantiated because there seem to be no books that easy or hard.

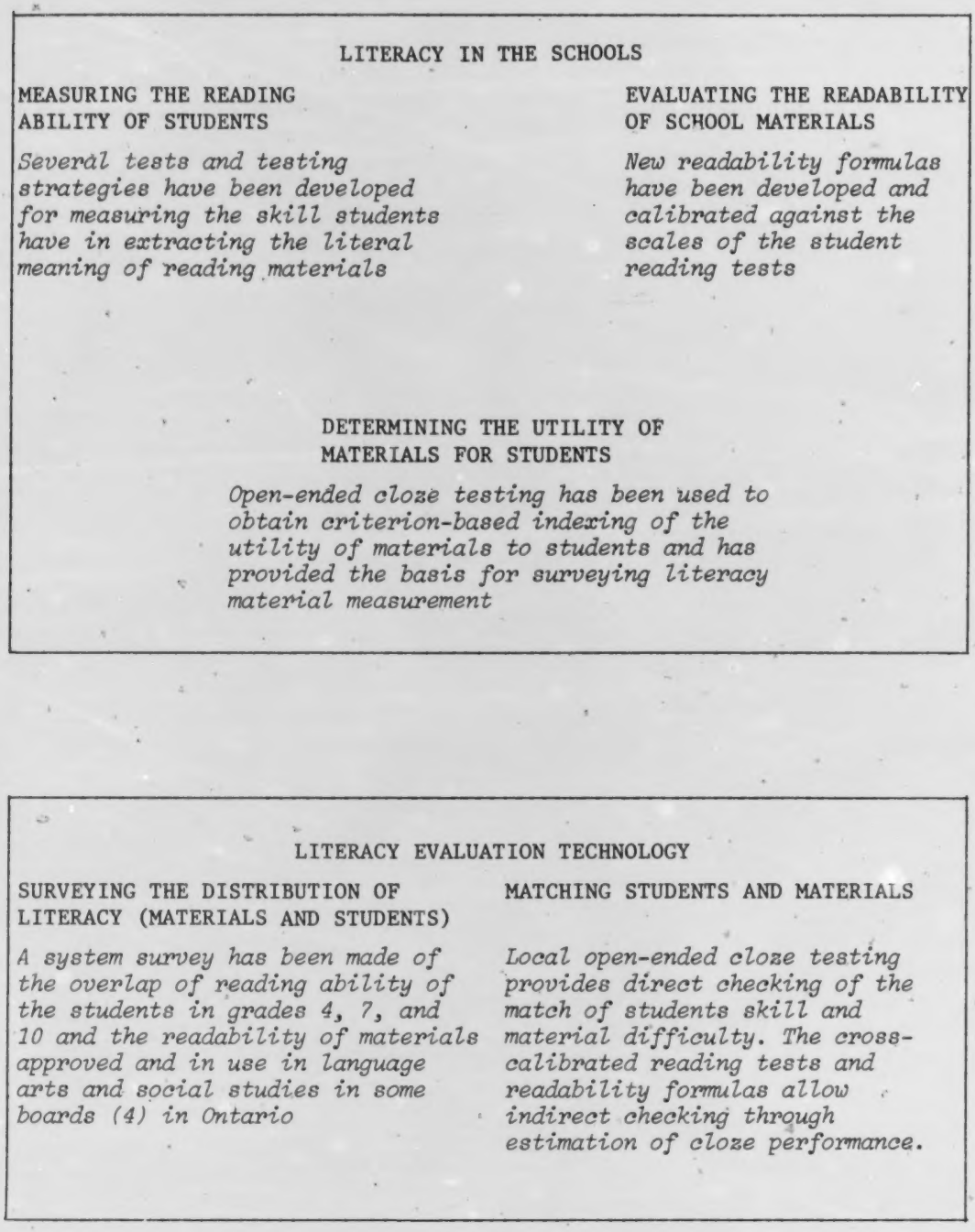
Table 20/ Relative Values of the Regression Slopes of Word Difficulty and Sentence Length

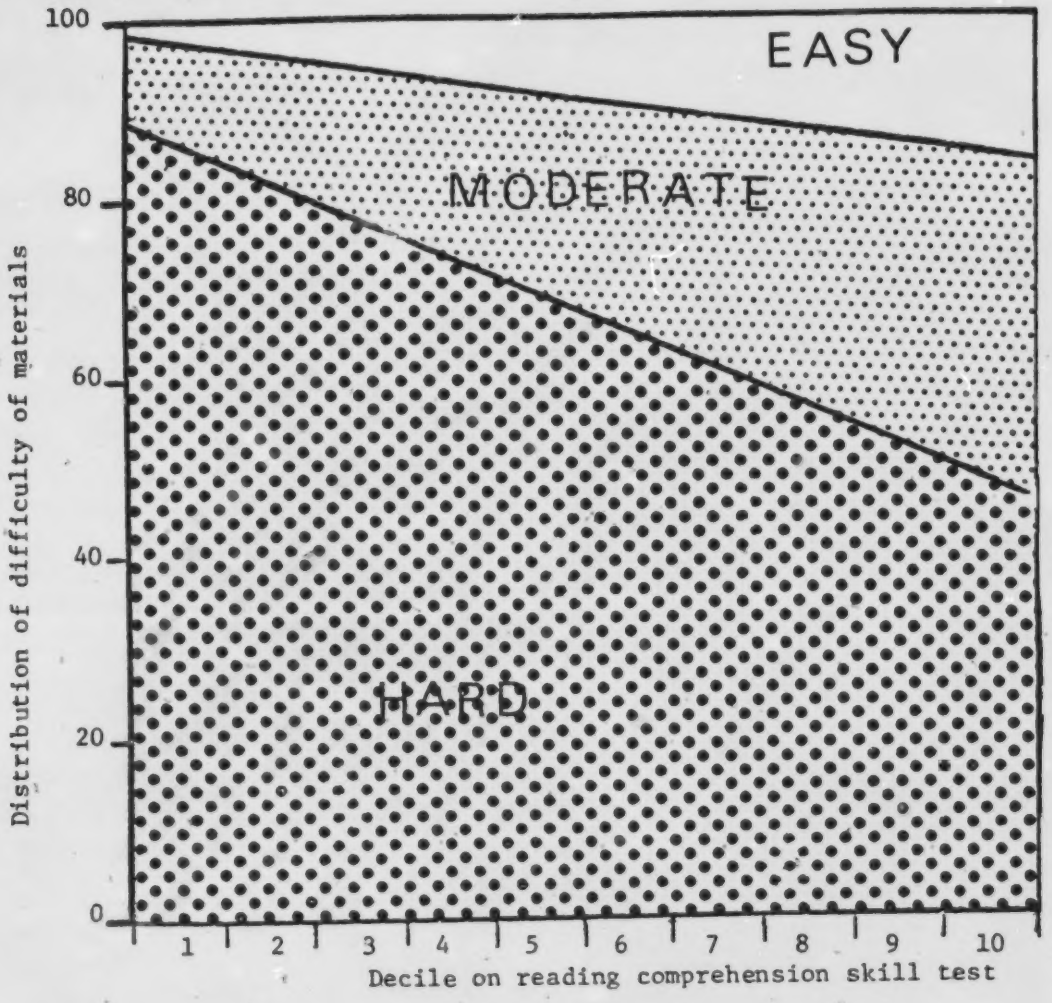
	Language Arts	Social Studies
Grade 4	2.30	.82
Grade 7	.97	.21
Grade 10	.44	.11

Note: These ratios come from the final calibration formulas and show that the relative impact of word difficulty compared to sentence length is higher for lower grades and for language arts.

FIGURES

Figure 1/ Overview of the Literacy Evaluation Project


LITERACY IN THE SCHOOLS
MEASURING THE READING ABILITY OF STUDENTS
Several tests and testing strategies have been developed for measuring the skill students have in extracting the literal meaning of reading materials
EVALUATING THE READABILITY OF SCHOOL MATERIALS
New readability formulas have been developed and calibrated against the scales of the student reading tests
DETERMINING THE UTILITY OF MATERIALS FOR STUDENTS
Open-ended cloze testing has been used to obtain criterion-based indexing of the utility of materials to students and has provided the basis for surveying literacy material measurement
LITERACY EVALUATION TECHNOLOGY
SURVEYING THE DISTRIBUTION OF LITERACY (MATERIALS AND STUDENTS)
A system survey has been made of the overlap of reading ability of the students in grades 4, 7, and 10 and the readability of materials approved and in use in language arts and social studies in some boards (4) in Ontario
MATCHING STUDENTS AND MATERIALS
Local open-ended cloze testing provides direct checking of the match of students skill and material difficulty. The cross-calibrated reading tests and readability formulas allow indirect checking through estimation of cloze performance.



Keys to Figures 2 to 7

2. Language Arts, Grade 4
3. Language Arts, Grade 7
4. Language Arts, Grade 10
5. Social Studies, Grade 4
6. Social Studies, Grade 7
7. Social Studies, Grade 10

Note: Each figure shows the distribution of the difficulty of current reading materials - what proportion of the materials are easy, moderate, and hard - for one grade level and subject area and are separated according to the reading comprehension skill of the potential reader. In all cases - most materials are hard for the poor reader, and even for the best readers - few materials are easy.

Figure 2 Distribution of Difficulty of Materials for Students of Different Levels of Reading Comprehension Skill, Language Arts, Grade 4

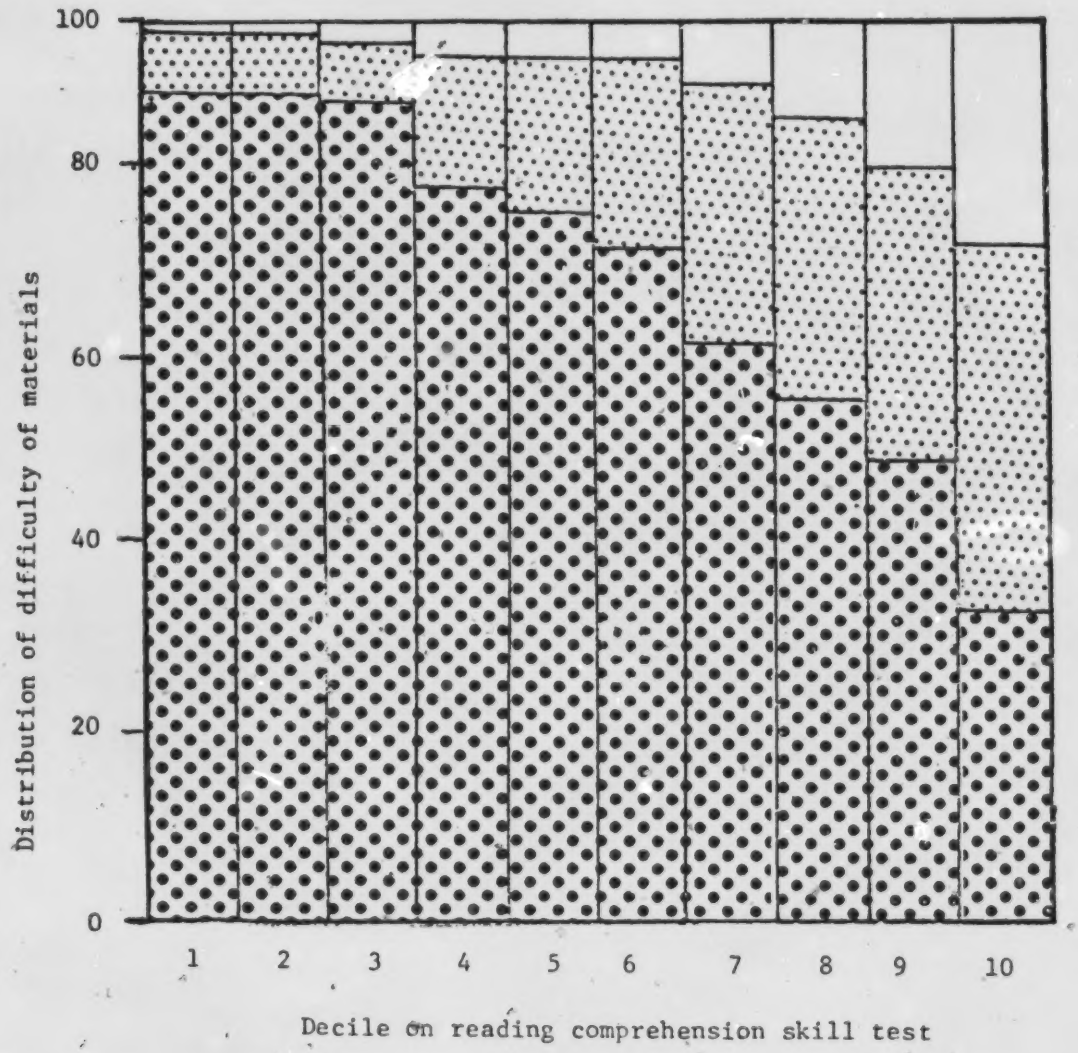


Figure 3 Distribution of Difficulty of Materials for Students of Different Levels of Reading Comprehension Skill, Language Arts, Grade 7

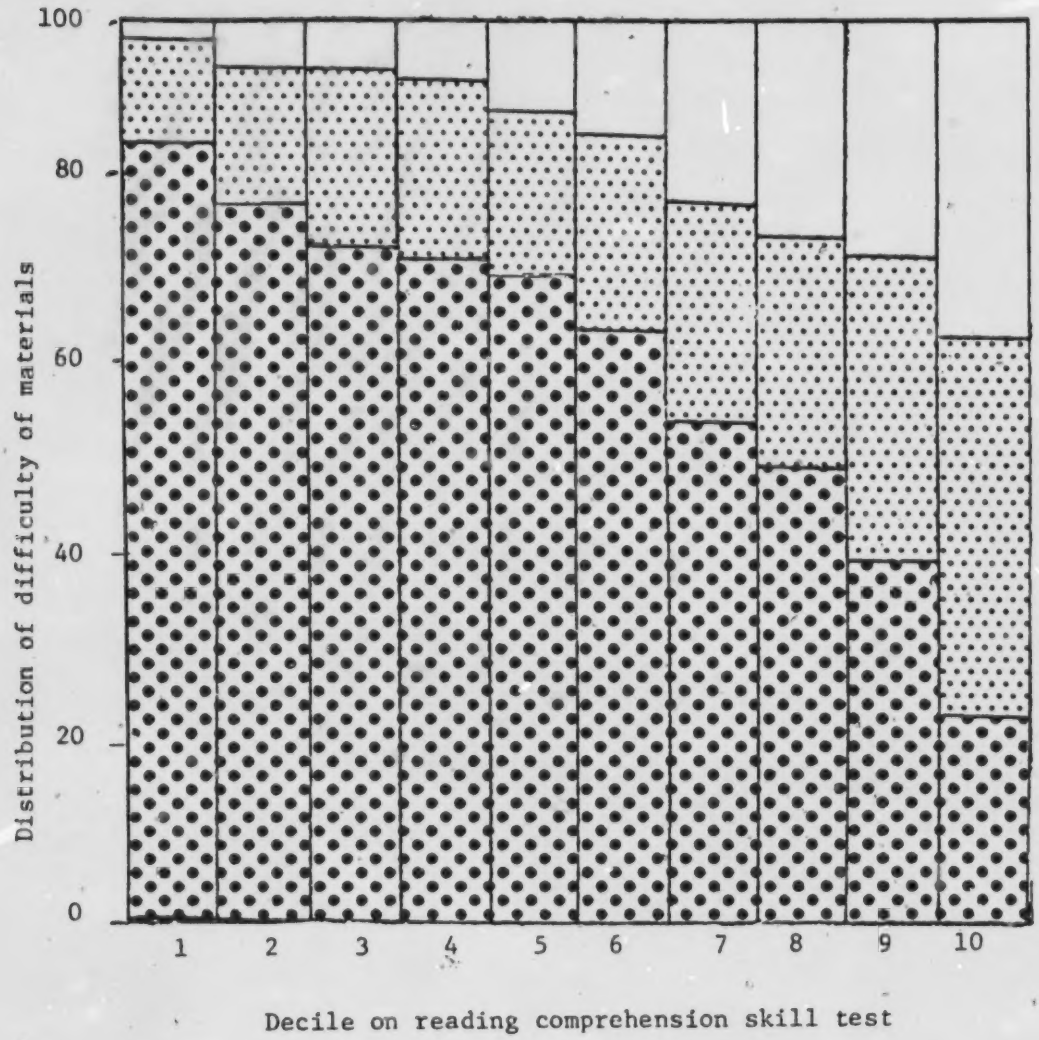


Figure 4 Distribution of Difficulty of Materials for Students of Different Levels of Reading Comprehension Skill, Language Arts, Grade 10

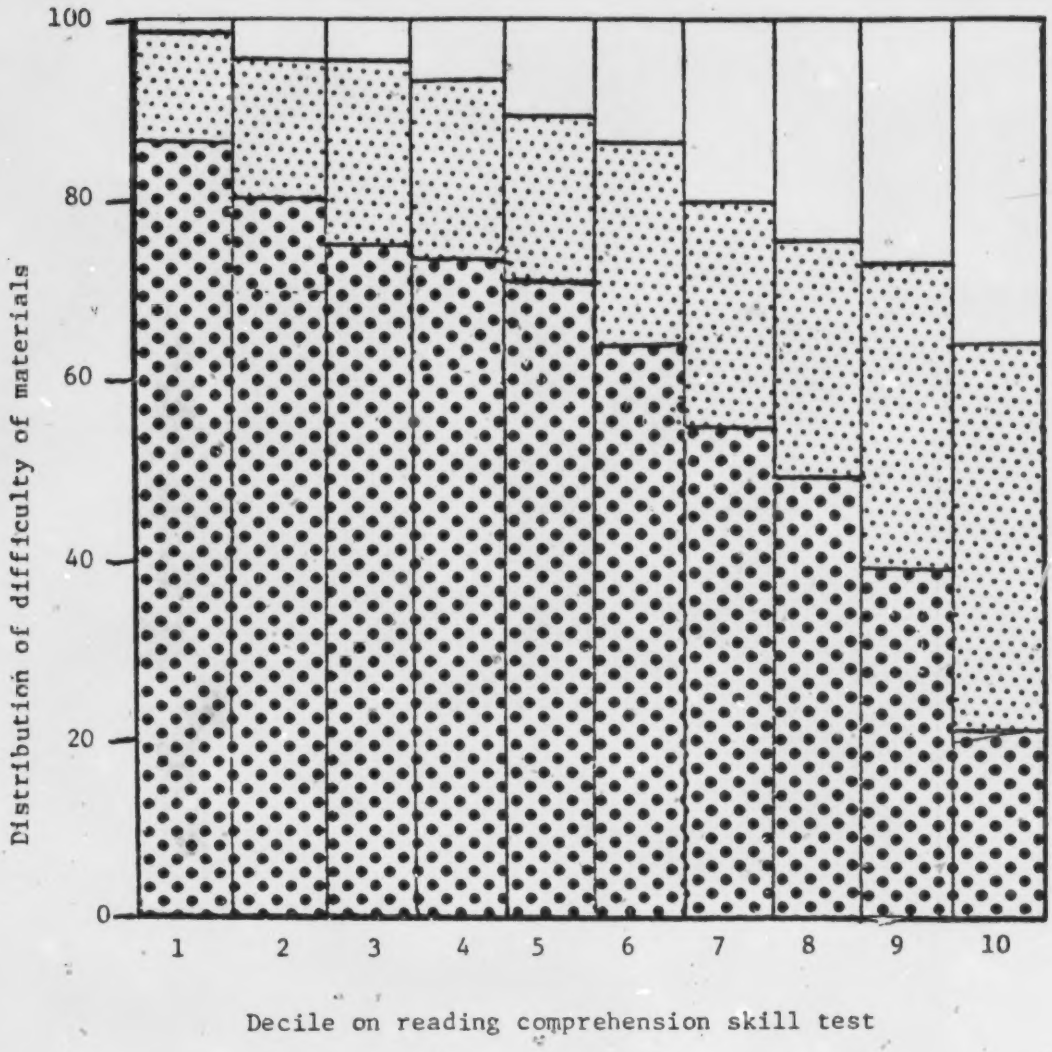


Figure 5 Distribution of Difficulty of Materials for Students of Different Levels of Reading Comprehension Skill, Social Studies, Grade 4

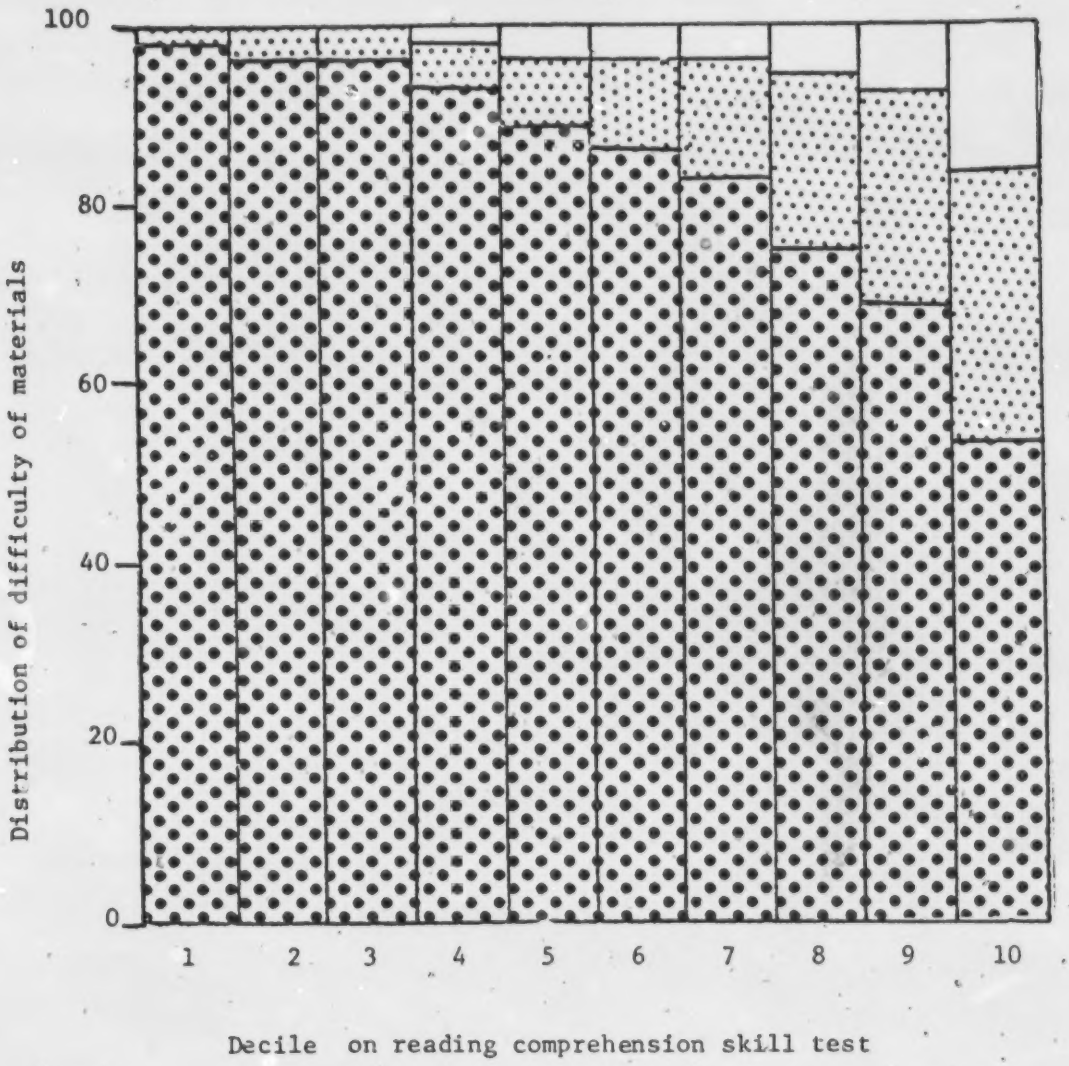


Figure 6 Distribution of Difficulty of Materials for Students of Different Levels of Reading Comprehension Skill, Social Studies, Grade 7

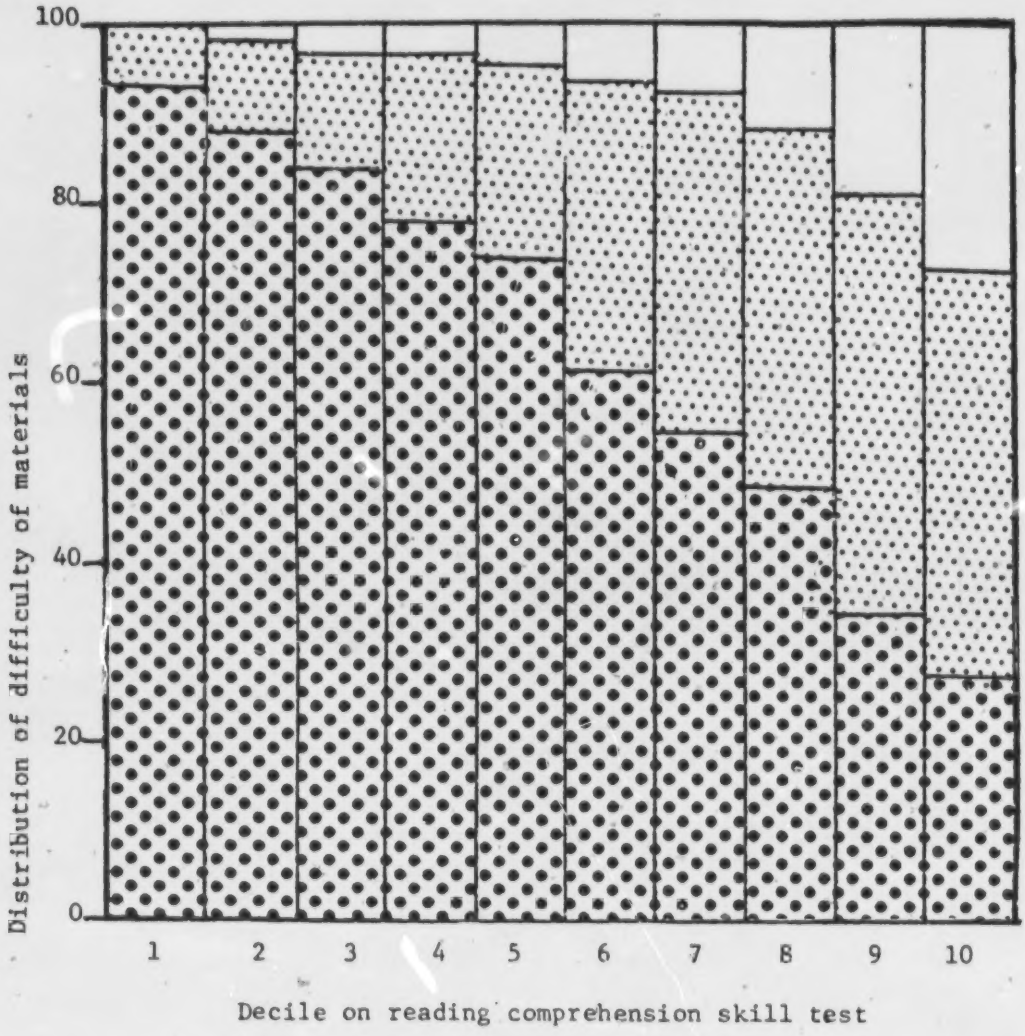
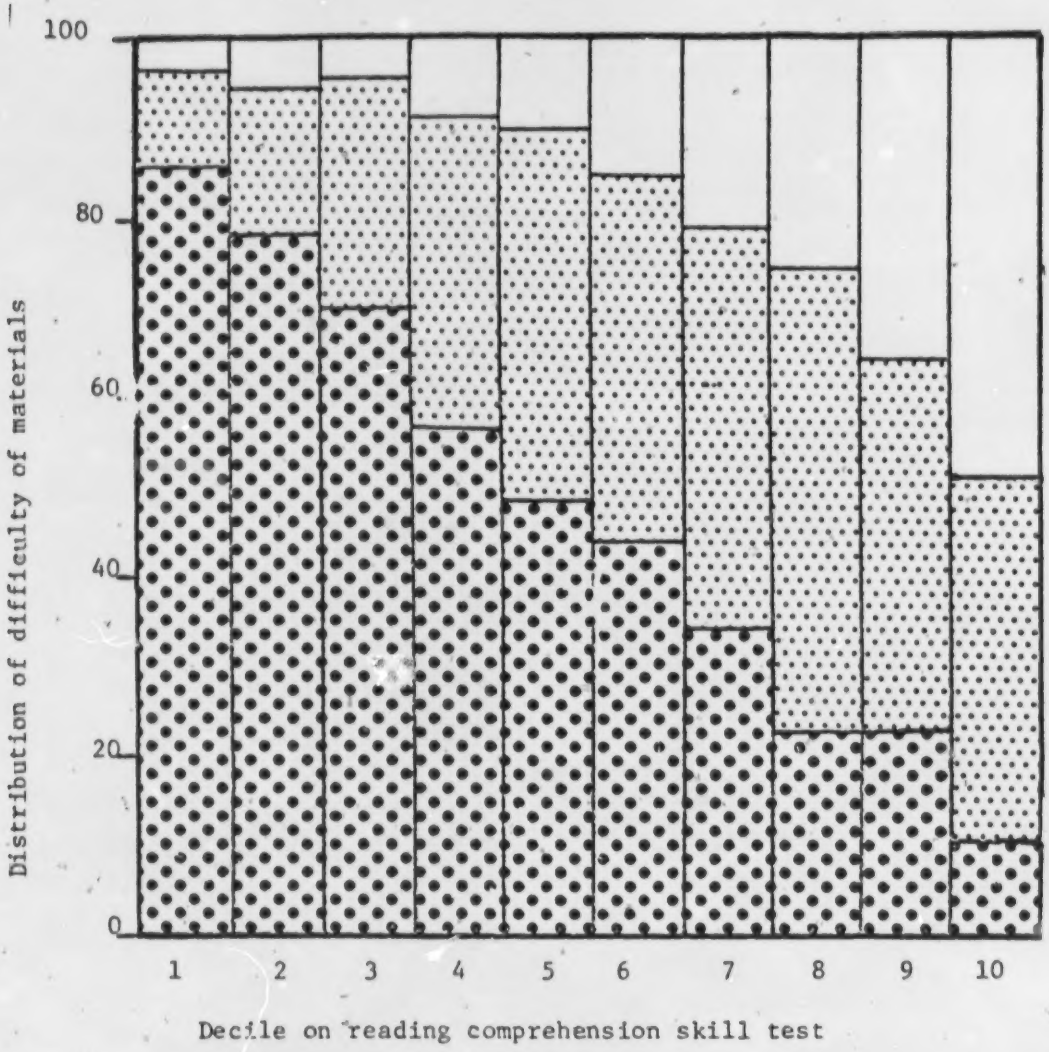
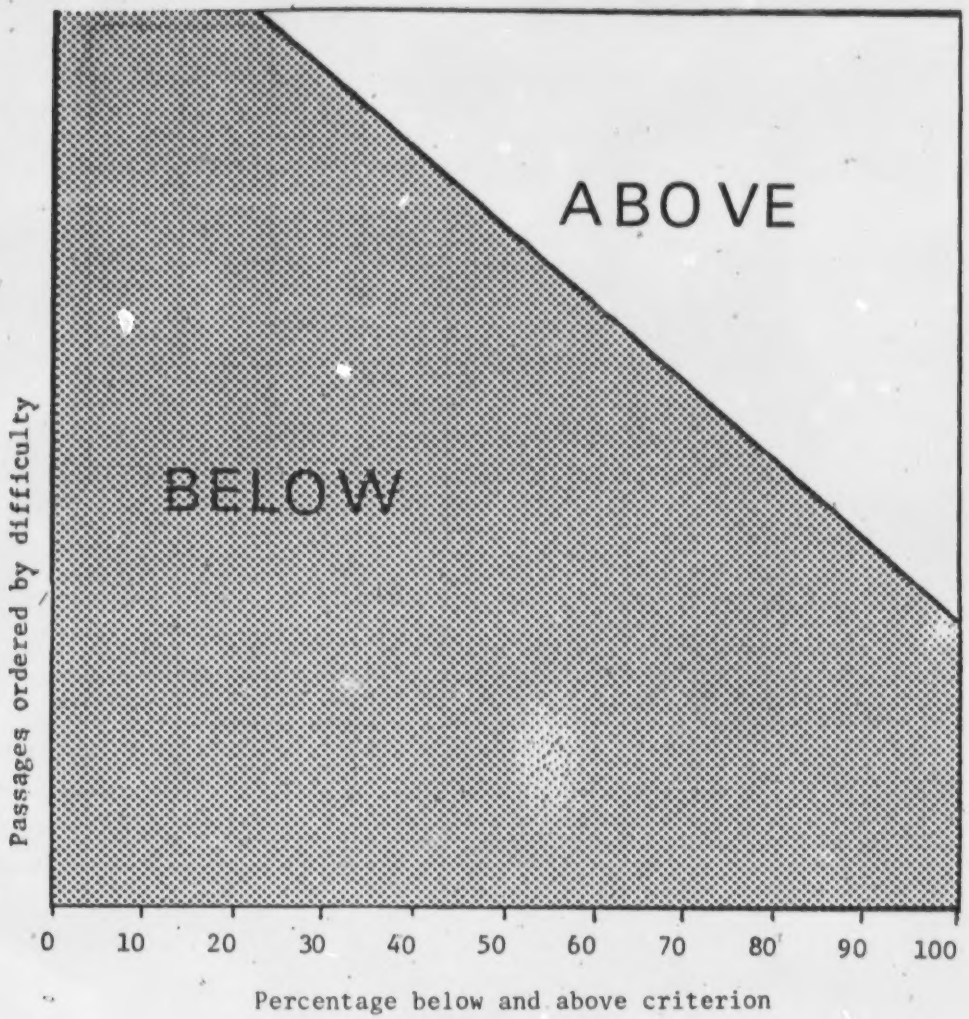


Figure 7 Distribution of Difficulty of Materials for Students of Different Levels of Reading Comprehension Skill, Social Studies, Grade 10





Keys to Figures 8 to 13

8 Language Arts, Grade 4
9 Language Arts, Grade 7
10 Language Arts, Grade 10
11 Social Studies, Grade 4
12 Social Studies, Grade 7
13 Social Studies, Grade 10

Note: Each figure above the literacy distribution of passages by students - what proportion of students are above and below criterion for each passage - for one grade level and subject area. The passages are ranked from easy to hard. For most passages, the majority of the students do not reach criterion.

Figure 8 Literacy Distribution of Passages by Students, Language Arts, Grade 4

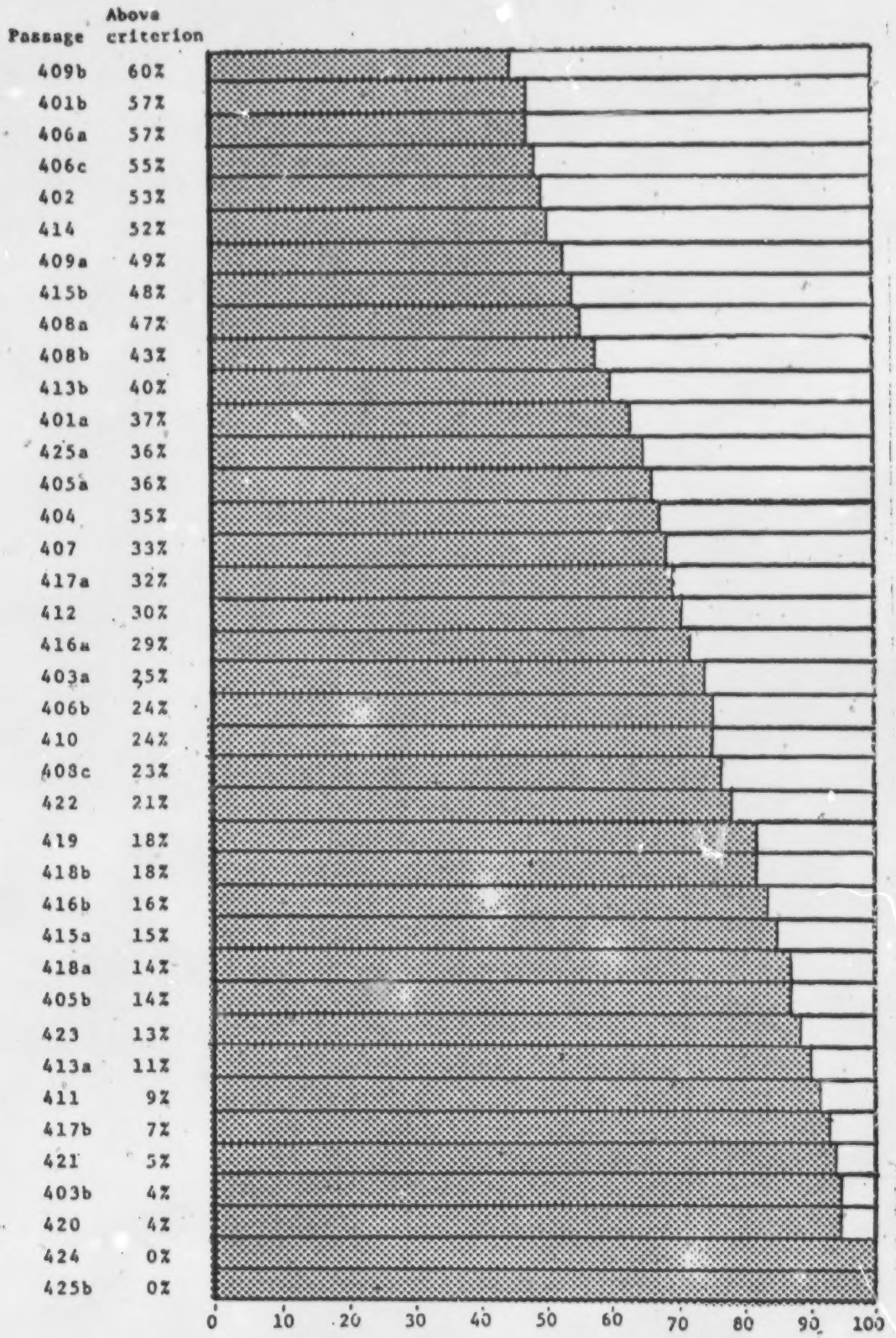


Figure 9 Literacy Distribution of Passages by Students, Language Arts, Grade 7

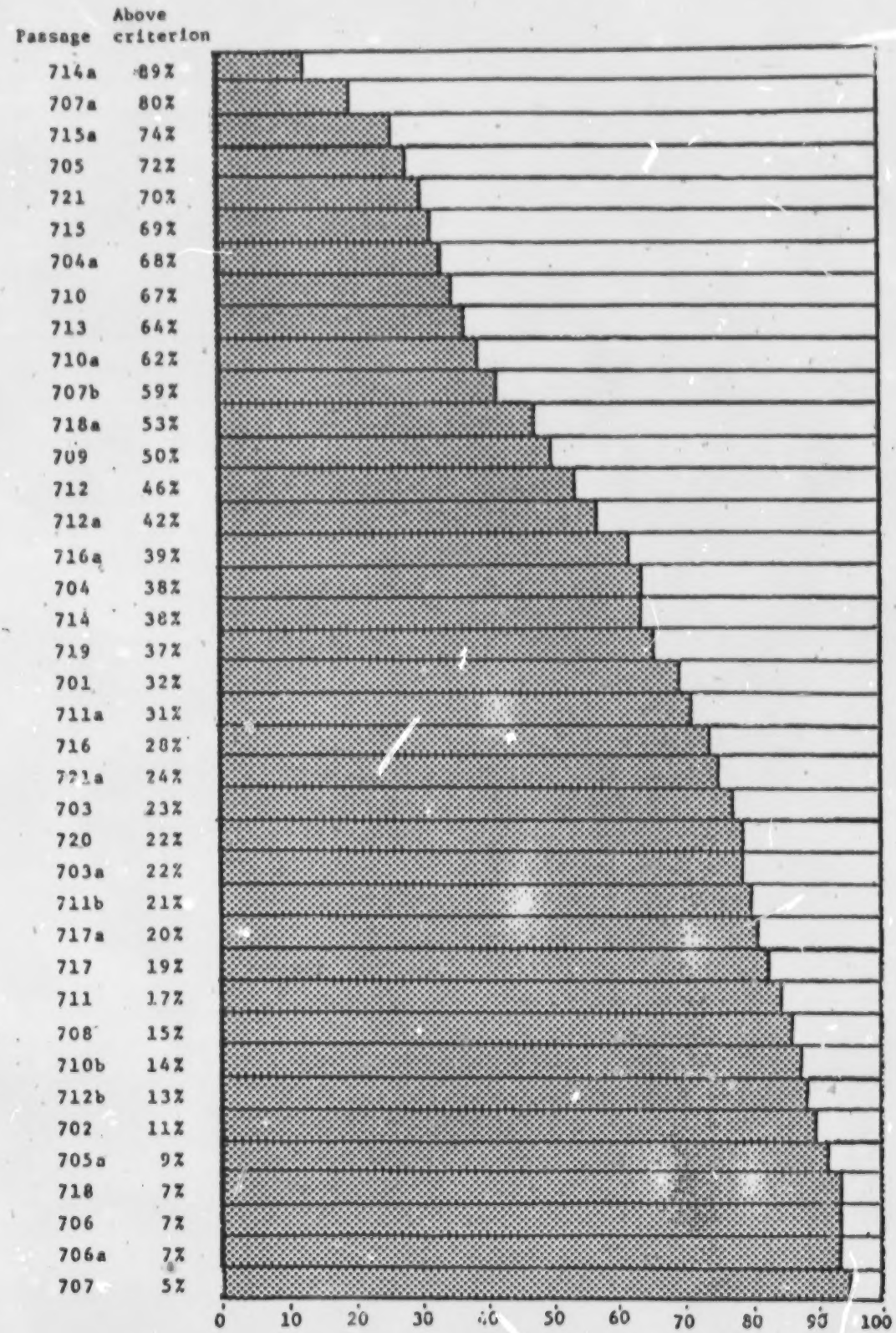


Figure 10 Literacy Distribution of Passages by Students, Language Arts, Grade 10

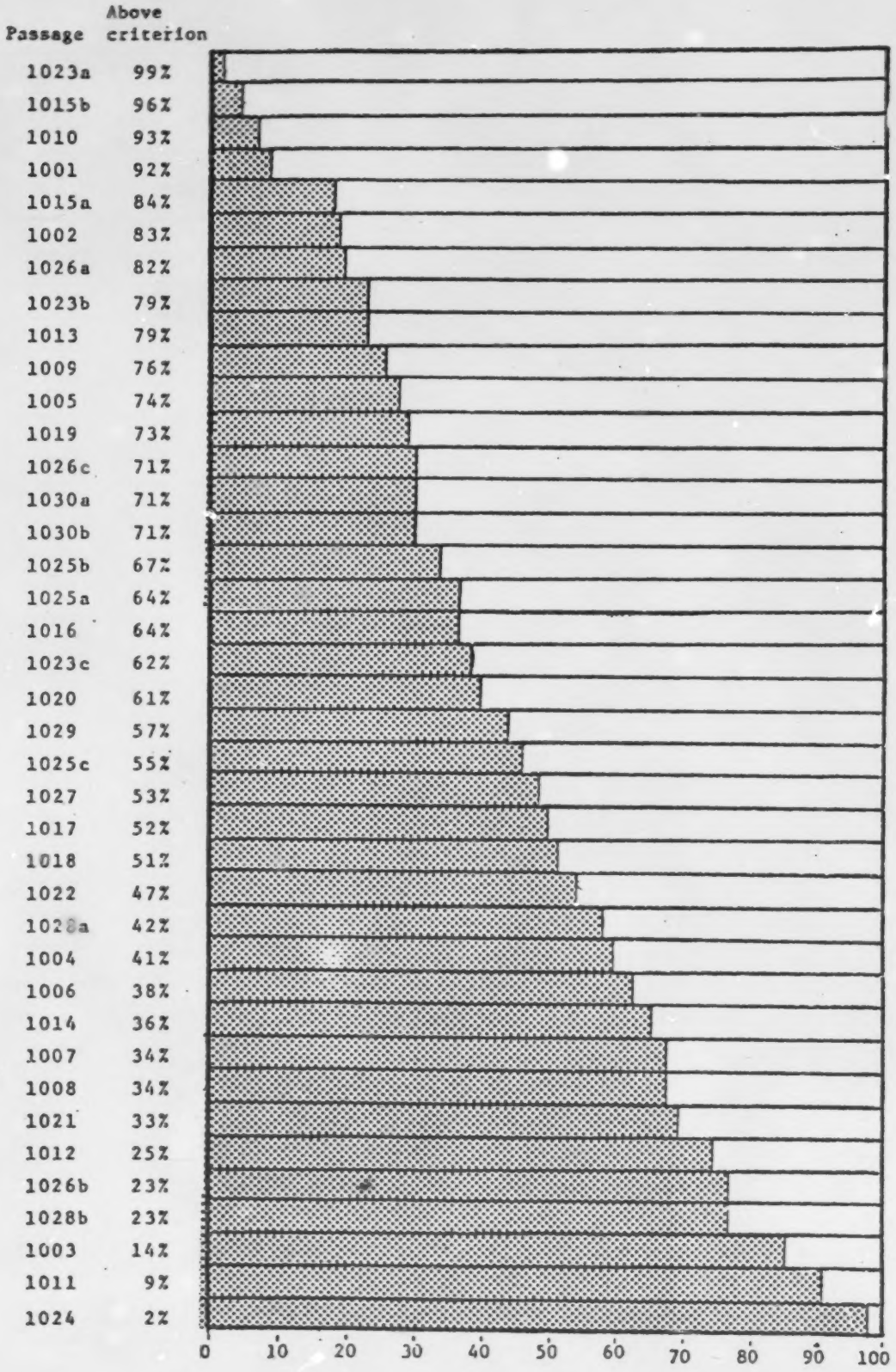


Figure 11 Literacy Distribution of Passages by Students, Social Studies, Grade 4

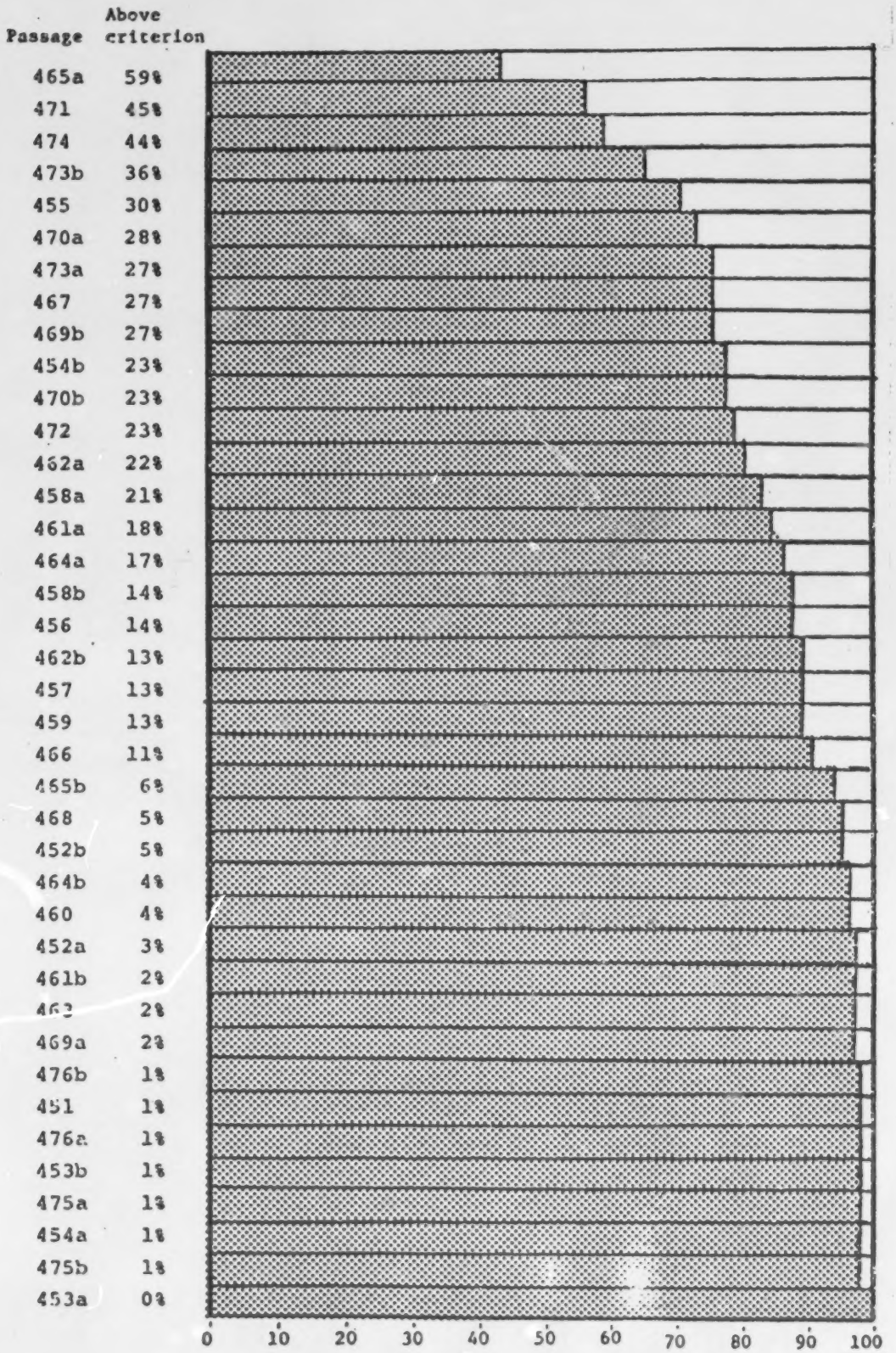


Figure 12 Literacy Distribution of Passages by Students, Social Studies, Grade 7

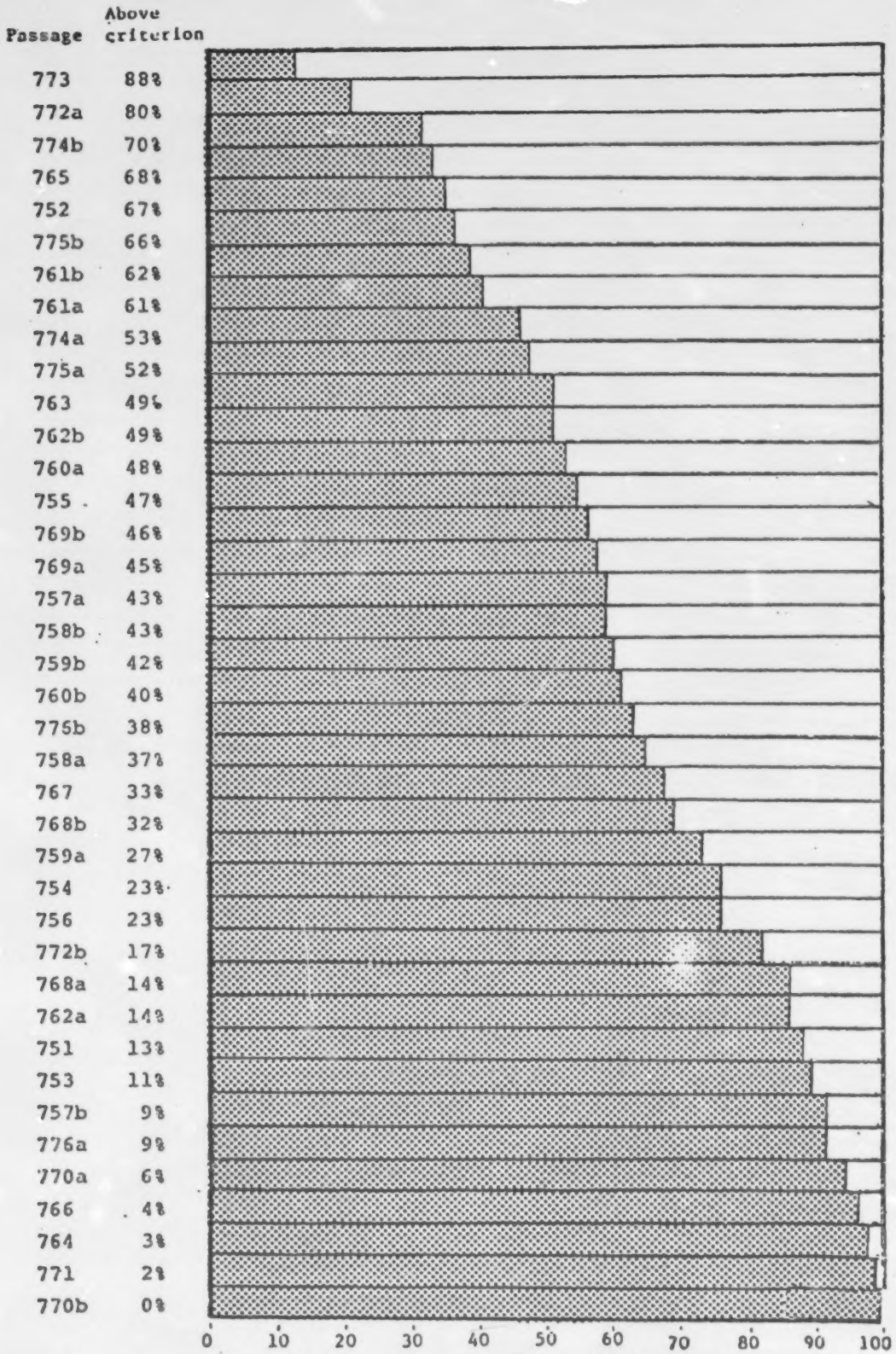


Figure 13 Literacy Distribution of Passages by Students, Social Studies, Grade 10

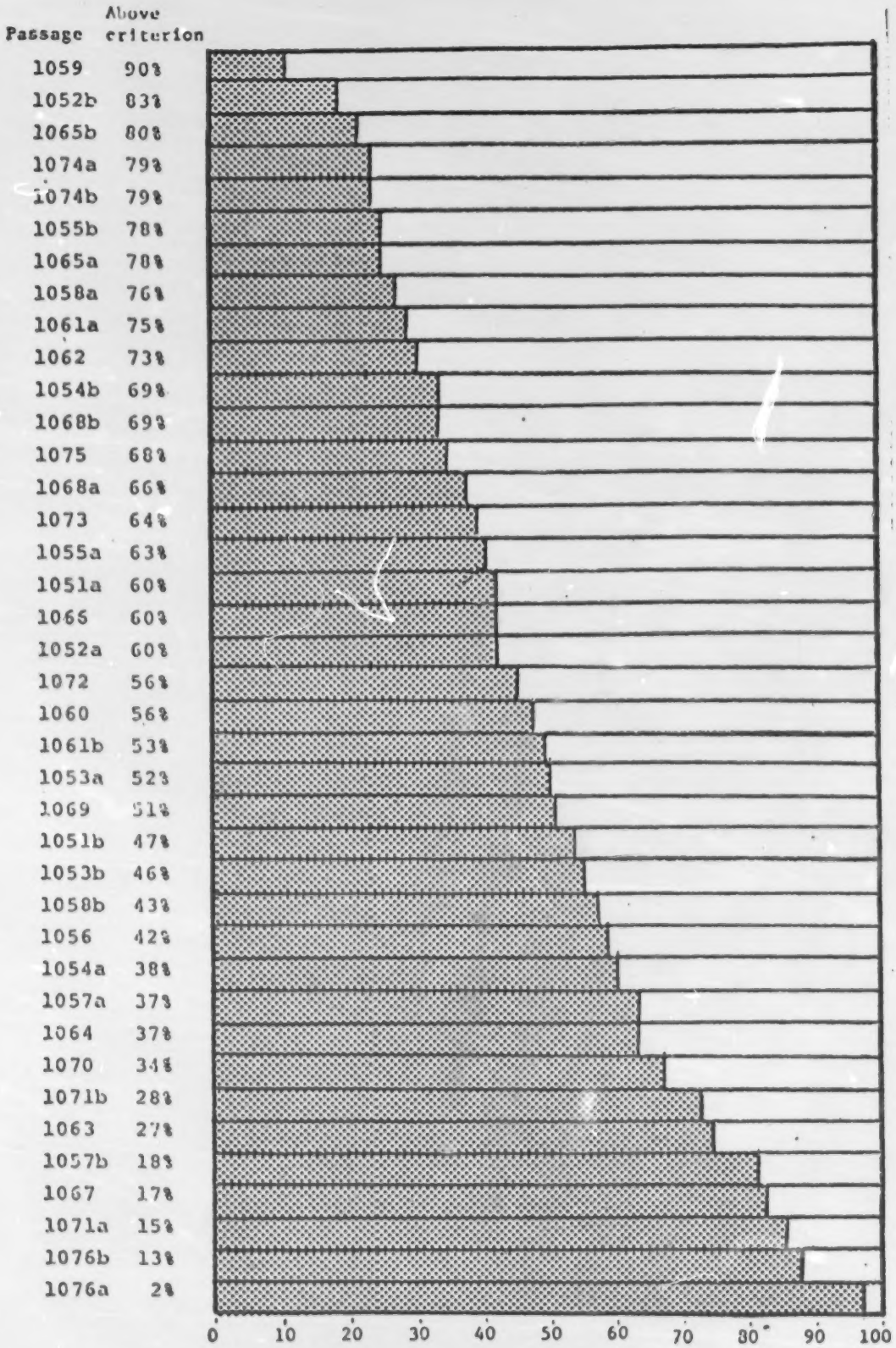
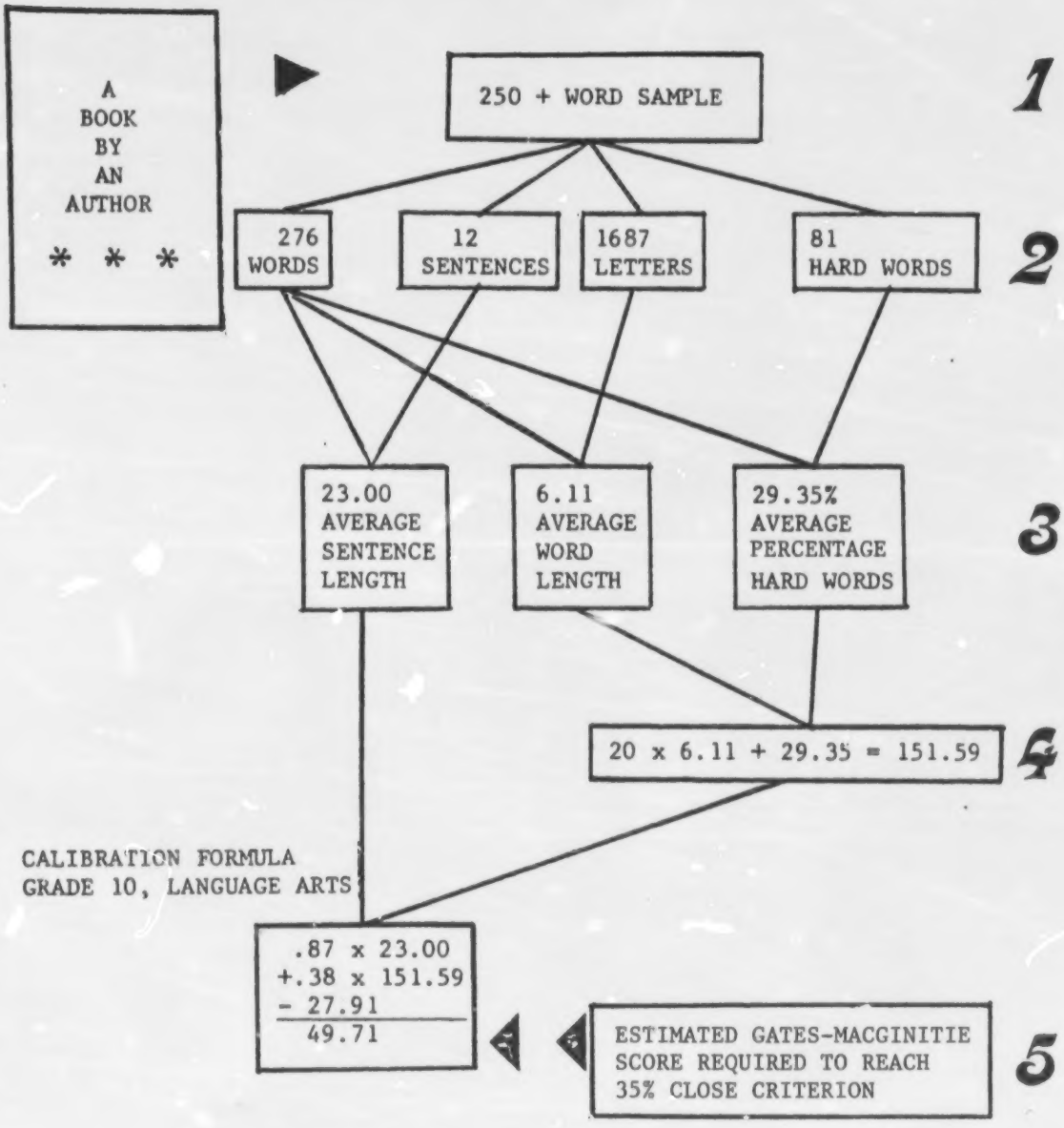


Figure 14 How to Find the Gates-MacGinitie Test Schore Estimated to Be Required for Literal Comprehension of a Passage in Grade 10, Language Arts



73

A1. LANGUAGE ARTS, GRADE 4

DETERMINATION

One day they began talking when they were pitching _____ in the barn. Danny _____ Judy all about the _____ and how he had _____ his chance.

"Was the _____ unfair?" asked Judy who _____ baseball, too.

"No," replied _____. "I made too many _____."

"Then looking sad won't _____," said Judy. "Why don't _____ practise pitching, catching, and _____ and make the coach _____ his mind?"

This sounded _____ a wonderful idea to _____. So right away after _____ he began work. First _____ nailed two thick sticks _____ an upright pole behind _____ barn. One stick was _____ as high as a _____ shoulder. The other was _____ as low as his _____. On the ground under _____, Danny placed a home-_____ cut from an old _____.

Hour after hour he _____. Sometimes Judy helped him, _____ more often he worked _____. When the balls hit _____ barn, Danny had to _____ them. His legs became _____ running after them. He _____ at the pitching until _____ balls went between the _____ sticks and across the _____ with great speed.

At _____ end of three weeks _____ hard work, Danny was _____ far better pitcher. It _____ time to make another _____. With his heart in _____ mouth, he asked the _____ again. The coach looked _____.

"The team's already picked," _____ said.

Those words should _____ been enough to stop _____. But not Danny! He _____ talking until he was _____ to join the afternoon _____.

THE TRANSATLANTIC CABLE

During three days of calm after the storm the _____ combed the bottom with _____ irons in an attempt _____ _____ recover the lost cable, _____ without success. Field received _____ bad news without turning _____ hair. "Let's try again!" _____ said. "Nothing ventured, nothing _____."

A year passed before _____ second expedition was ready _____ sea. In the meantime _____ new and different plan _____ attack was made. The _____ and the Niagara met _____ mid ocean where the _____ of the twelve-hundred-_____ cables were spliced together. _____ everything was in readiness, _____ ships sailed together. One _____ the east, the other _____ the west, letting out _____ spliced cable as they _____ gingerly.

Now that the _____ were joined by wire, _____ were able to communicate _____ each other. As if _____ its former anger, the _____ was kindly. With a _____ sea, warm sunshine, and _____ nights, work went on _____ the clock. Hopes were _____ with each passing hour.

_____ after dawn one fine _____ a joyful cry came _____ the crow's- nest on _____ Agamemnon. "Land ho! Land _____!" A cheer went up _____ the crew and a _____ blast was sounded by _____ Agamemnon's whistle. A few _____ later, the Niagara reported _____ telegraph over the cable _____ it, too, was in _____ of its destination, which _____ this occasion was the _____ coast.

Before sunset that _____ in 1858 both ends _____ the cable had been _____ on shore and connected _____ their respective stations in Canada and Ireland.

THE BUTLER

And Joseph went about testing each of the gilded _____ to see whether it _____ moved since he had _____ placed it. Joseph habitually _____ at furniture, expecting it _____ be impertinent, mischievous, or _____. In a world where _____ Orden was the leader _____ men, Joseph was the _____ of furniture, silver, and _____. Joseph was elderly and _____ and serious, and his _____ was so complicated that _____ a profound man would _____ him to be simple. _____ saw nothing amazing about _____ Winter's rolling thumbs; in _____ he found them irritating. _____ suspected that something pretty _____ was happening, what with _____ soldiers in the town _____ the local army killed _____ captured. Sooner or later _____ would have to get _____ opinion about it all. _____ wanted no levity, no _____ thumbs, no nonsense from _____. Doctor Winter moved his _____ a few inches from _____ appointed place and Joseph _____ impatiently for the moment _____ he could put it _____ again.

Doctor Winter repeated, "_____ o'clock, and they'll be _____ then, too. A time-_____ people, Joseph."

And Joseph _____, without listening, "Yes sir."

"_____ and machines."

"Yes sir."

"_____ hurry toward their destiny _____ though it would not _____. They push the rolling _____ along with their shoulders."

_____ Joseph said, "Quite right, _____," simply because he was _____ tired of saying, "Yes _____."

Joseph did not approve _____ this line of conversation, _____ it did not help _____ to have an opinion _____ anything.

A4. SOCIAL STUDIES, GRADE 4

PIONEER SCHOOLS

The class would be made up of students of _____ ages and grades, since _____ one teacher would be _____ for the instruction of _____ in the neighbourhood. The _____ exceptions might be the _____ older children from quite _____ families who would be _____ away to live and _____ at one of the _____ academies. Many of the _____ in the local school _____ through the grades very _____, since they would attend _____ only when they were _____ needed to help at _____. Many families during their _____ pioneer years needed every _____ to be turned to _____ a homestead out of _____ forest. Most children attended _____ over the course of _____ or six years, but _____ spent much less time _____ studying. It may have _____ just as well that _____ schoolmasters confined their instruction _____ the "Three R's, Readin', _____, an' 'rithmetic."

Although there _____ some teachers who were _____-educated and who passed _____ some of their knowledge _____ languages, history and science _____ their pupils, most were _____ quite thankful to have _____ teach only the rudiments, _____ is, the simple skills. _____ average schoolmaster was not _____ trained for the job _____ may have had little _____ learning than his older _____. As a general rule, _____ could not expect to _____ well paid by communities _____ were themselves struggling to _____. In addition to a _____ salary, which might total _____ dollars per year for _____ pupil, the teacher could _____ to receive free board in a local home.

A PROVISIONAL GOVERNMENT IS FORMED

At the end of the second day Riel admitted _____ had come to the _____ "with fear". "We are _____ yet enemies," he said _____ his audience, "but we _____ very near being so. _____ soon as we understood _____ other, we joined in _____ what our English fellow-_____ in common with us _____ to be our just _____. I am not afraid _____ say 'our rights' for _____ all have rights. We _____ no half-rights, mind _____, but all the rights _____ are entitled to. Those _____ will be set forth _____ our representatives, and what _____ more, gentlemen, we will _____ them." Several times during _____ speech the crowd interrupted _____ loud cheers and the _____ broke up with the _____ that the crisis at _____ River was over.

In _____ last speech Riel had _____ that French and English _____ give consideration to Smith's _____. The spirit of goodwill _____ had flowed through the _____ with Smith did not _____ over into the representatives' _____, however. While the representatives _____ discussing a new Bill _____ Rights, it became clear _____ Riel could no longer _____ his own way unopposed. _____ never accepted criticism or _____ easily and when he _____ that some of the _____ representatives were not prepared _____ accept several of his _____, he gave vent to _____ outbursts of temper.

Smith _____ been invited to attend _____ meetings. As each item _____ the Bill of Rights _____ discussed, Riel asked him _____ he could guarantee that the Canadian government would agree.

A6. SOCIAL STUDIES, GRADE 10

NATURAL VEGETATION

Between the Arctic tundra and the coniferous forest, there _____ a transition zone of _____ coniferous trees and shrubs _____ with tundra vegetation. This _____ is called the Subarctic. _____ most common coniferous trees _____ this transition zone are _____ and larch. There are _____ some white birch, one _____ the few deciduous trees _____ can survive such a _____ climate.

Because of the _____ growing season and the _____ amount of precipitation, the _____ of trees is very _____. As a result, most _____ the trees are small, _____, and of no use _____ man. For this reason, _____ area is known as _____ "land of the little _____." The rocky areas, having _____ trees and very little _____ vegetation, are called barrens.

_____ Subarctic region has many _____ and swamps. Swampy areas, _____ bogs, that have filled _____ with peat are called _____. Spruce and tamarack trees _____ densely around the edges _____ these muskegs. Towards the _____, there are many shrubs _____ cranberries. These muskegs have _____ a transportation barrier to _____ travel in the north. _____ in winter, when they _____ frozen over, is it _____ to cross the muskegs.

_____ of the wildlife of _____ Arctic is also found _____ the Subarctic region. The _____, lemming, and great snowy _____, are all common. There _____ also very many animals _____ the Subarctic that are _____ associated with the Coniferous _____ region to the south. _____ summary, it may be _____ that the Subarctic is _____ transitional zone of wildlife;

BIBLIOGRAPHY

Agin, A.P. Recent research in reading, Education Digest, September, 1975, 41, 45-48.

Allen, P.D. (Ed.) Findings of research in miscue analysis: Classroom implications. Illinois: ERIC Clearing House on Reading and Communication Skills, 1976.

Anderson, J. Research in readability for the classroom teacher. The Journal of Reading, February, 1967, X, 291-297.

Ayrer, James. Problems in the development of a test of functional literacy. Journal of Reading, 1977, 20, 697-705.

Babcock, Natalie. Cloze procedure and the affective domain. Paper presented at the 8th Annual Conference of the Western College Reading Association, Anaheim, California, March 20-22, 1975.

Bailey L.J. Cloze procedures- A procedure for evaluating instructional materials. Journal of Industrial Teacher Education, Summer, 1969, 6(4), 57-60.

Ball, I.L., & Williamson, H.J. How readable is literature for children. V.I.E.R. Bulletin, 1973, 31, 14-20.

Beil, Drake. The emperor's new cloze. The Journal of Reading, April, 1977, 601-604.

Bickley A.C. The cloze procedure: A conspectus. Journal of Reading Behavior, Summer, 1970, 2(3), 234-249.

Borich, G.D. (Ed.) Evaluating educational programs and products. New Jersey: Educational Technology Publications, 1974.

Bormuth J.R. Relationships between selected language variables and comprehension ability and difficulty. Los Angeles: University of California, 1964. (Eric Document ED003 325)

Bormuth J.R. The effectiveness of current procedures for teaching reading comprehension. Paper presented at the National Council of Teachers of English meeting, Milwaukee, November 30, 1968.

Bormuth J.R. Factor validity of cloze test as measures of reading comprehension. Reading Research Quarterly, Spring ,1969, 4, 358-365.

Bormuth J.R. The implications and use of cloze procedures in the evaluation of instructional programs. Los Angeles: University of California Center for the Study of Evaluation of Instructional Programs, No. 3, 1967. (Eric Document ED 012 674)

Bormuth J.R. Readability - A new approach. Reading Research Quarterly, 1966, 1, 79-131.

Bormuth J.R. Cloze test readability: Criterion reference scores. Journal of Educational Measurement, 1968, 5, 189-196.

Bormuth, J.R. Optimum sample size and cloze test length in readability measurement. Journal of Educational Measurement, June, 1965, 2, 111-116.

Bormuth J.R. Cloze as a measure of readability. In International Reading Association Conference Proceedings, 1963, 8, 131-134.

Bormuth, J. R. Illiteracy in the suburbs. Unpublished paper, University of Chicago, 1970a.

Bormuth, J. R. On the theory of achievement test items. Chicago, Illinois: University of Chicago Press, 1970b.

Bormuth, J. R. Development of standards of readability: Toward a rational criterion of passage performance. Final Report, Project No. 9-0237. Bureau of Research, U.S. Office of Education, 1971.

Bormuth, J.R. Readability formulas and the literacy production function. University of Chicago, Research project funded by U.S. Office of Education and the Reading Research Fund, 1977.

Bormuth, J.R. Value and volume of literacy. Visible Language, 1978, 12(2), 118-26.

Bortnick, R., & Lopardo, G.S. The case for cloze in the classroom. Paper presented to the International Reading Association, New Orleans, May, 1974.

Bortnick, R., & Lopardo, G.S. The cloze procedure: A multi-purpose classroom tool. Reading Improvement, 1976, 13(2), 113-117.

Boruch, R.F., & DeGracie, J.S. The use of regression discontinuity model with criterion-referenced testing in the evaluation of compensatory education. Washington, D.C.: National Institute of Education, 1977. (ERIC Document ED 135 869)

Boulton, M. Reading for real life. London: Macmillan Education Ltd., 1971.

Breneman, Beth. Reaction: The cloze and the composition process. Paper presented at the 65th Annual Meeting of the National Council of Teachers of English, San Diego, California, November 26-29, 1975.

Boyce M. W. Some difficulties in using cloze procedures to access readability. M.ED. Thesis, University of Melbourne, 1974. (ERIC Document Reproduction Service No. ED 110 921)

Carroll, J.B. Measurement of meaning. Language, 1959, 35, 58-77.

Caratens, P.W., & McKeag, R.A. Cloze test scores for assessing reading comprehension: Some considerations for interpreting data. Reading Improvement, 1975, 12(1), 8-10.

Chester, Robert, & Dulin, Ken. Three approaches to the measurement of secondary school students' attitudes towards books and reading. Research in the Teaching of English, 1977, 2(2), 193-200.

Chomsky, N. Syntactic structures. The Hague: Mouton, 1957.

Chomsky, N. Aspects of the theory of syntax. Cambridge, Mass: M.I.T. Press, 1965.

Cohen, J.H. The effect of content area material on cloze test performance. The Journal of Reading, December, 1975, 19(3), 247-250.

Cole, G. The chains of functional illiteracy. The Education Digest, December, 1977, 10-13.

Cooper, C.L., & Petrocky, A.R. Reading strategies and teaching implications for secondary schools from the psycholinguistic model of the reading process. The High School Journal, November, 1975, 59(2), 91-102.

Cooper, C.L., & Petrosky, A.R. A psycholinguistic view of the fluent reading process. The Journal of Reading, December, 1976, 20(3), 184-207.

Cortes, Carlos E., & Campbell, Leon G. Forward to basics: Teaching media literacy through the social studies. The Link, November,1977, n.p.

Cromer, R.F. Children are nice to understand. British Journal of Psychology, 1970, 61(3), 397-408.

Culhane, J.W. Comparable cloze and multiple choice comprehension test scores. The Journal of Reading, December, 1969, 13(3), 193-198.

Culhane, J.W. Cloze procedure and comprehension. Reading Teaching, February,1970,23, 725-755.

Dale, Edgar, & Chall, Jeanne S. A formula for predicting readability. Educational Research Bulletin, 1948, 27(5), 11-20, 37-54.

Dale, E., & Seels, B. Readability and reading - An annotated bibliography. London: International Reading Association, 1966.

Davis, F.B. (Ed.) The literature of research in reading with emphasis on models. New Jersey: Graduate School of Education, Rutgers, 1971.

Davis, Jerry. Improving reading and the teaching of science. The Clearing House, 1977, 50, 391-393.

deVilliers, J.G., & deVilliers, P.A. Development of the use of word order in comprehension. Journal of Psycholinguistic Research, 1973, 2(4), 331-341.

Dishner, Ernest K. The cloze procedure: A valuable tool for content teachers. Paper presented at the Annual Meeting of the National Reading Conference, 1973.

Dubois, S.V.C. (Ed.) Communications: A focus for partnership in education. Toronto: Ontario Association for Curriculum Development, 1976.

Dulin, K.L. Measuring the difficulty of reading materials. Reading Improvement, Spring, 1971, 8(1), 3-6.

Estes, T.H. A scale to measure attitudes toward reading. The Journal of Reading, November, 1971, 135-138.

Evaluation of Project Components: 1974-75, Maui District Office. Social Welfare Development and Research Center, Hawaii University, Honolulu, July, 1975.

Farr, Roger, & Weintraub, Sam. Measurement in reading research. Reading Research Quarterly, Fall, 1970, 6, 3-4.

Feely, T.M. How to match reading materials to student reading levels: The cloze and the maze. Social Studies:Nov./Dec., 1975, 66(6), 249-252.

Finn, Patrick J. Word frequency, information theory, and cloze performance: A transfer feature theory of processing in reading. Reading Research Quarterly 1977-1978, 4, 508-537.

Flesch, R. Estimating the comprehension difficulty of magazine articles. Journal of General Psychology, 1943, 28, 63-80.

Flesch, R. A new readability yardstick. Journal of Applied Psychology, 1948, 32, 221-233.

Flesch, R. Measuring the level of abstraction. Journal of Applied Psychology, 1950, 34, 384-390.

Fram, Ralph D. A review of the literature related to the cloze procedure. M. Ed. Thesis, Boston University School of Education, 1972.

Froese, V. Cloze readability versus the Dale-Chall formula. Research in Education, 1971. (ERIC Document Reproduction Service No. ED 051 975)

Fry, E.B. Developing a word list for remedial reading. In L.M. Schell and P.C. Burns (Eds.), Remedial reading: An anthology of sources. Boston: Allyn and Bacon,1968. Pp.361-66.

Fry, E.B. Reading instruction for classroom and clinic. New York: McGraw-Hill, 1972.

Ganier, A.S. The post oral reading cloze test: Does it really work? Reading World, October, 1976, 16(1), 21-27.

Gilliland, J. Readability. London: University of London Press, 1972.

Glover, Michael. Quickie tests for reading: Fast, simple tests to determine reading level. New Mexico State Department of Education, Santa Fe Division of Instructional Services, 1974.

Gold, J. (Ed.) In the name of language! Toronto: Macmillan, 1975.

Gomberg, Adeline W. Freeing children to take a chance. Reading Teacher, 1976, 29(5), 455-457.

Goodman, K.S. A linguistic study of cues and miscues in reading. Elementary English, 1965, 42(6), 639-645.

Goodman, K.S. Reading a psycholinguistic guessing game. In Harry Singer & Robert Ruddell (Eds.), Theoretical models and processes of reading. Newark, Delaware: International Reading Association, 1970. Pp. 259-271.

Goodman, K.S. Psycholinguistic universals in the reading process. In F. Smith (Ed.), Psycholinguistics and reading. Toronto: Holt,Rinehart and Winston, 1973. Pp. 21-27.

Goodman, K.S., & Fleming, J. Psycholinguistics and the teaching of reading. Newart, Delaware: International Reading Association, 1968.

Gray, W.S., & Leary, B.A. What makes a book readable? Chicago: University of Chicago Press, 1935.

Greathouse, L.J., & Neal, B.J., Letter cloze will conquer contractions. The Reading Teacher, November, 1976, 30(2), 173-176.

Gromberg, A.W. Freeing children to take a chance. The Reading Teacher, February, 1976, 29(5), 455-457.

Grove, M.K. Using the cloze procedure in a first grade classroom. The Reading Teacher, October, 1975, 29(1), 36-38.

Gunning, R. The technique of clear writing. New York: McGraw-Hill, 1952.

Hafner, L.E. Cloze procedure. The Journal of Reading, May, 1966, 9(6), 415-21.

Hansen, L.H., & Hesse, K.D. An interim report of results of the pilot assessment of reading literacy. Madison: Wisconsin State Department of Public Instruction, April, 1972. (ERIC Document Reproduction Service No. ED 079 704)

Hansen, L.H., & Hesse, K.D. A criterion-referenced assessment of reading literacy using the cloze procedure. Illinois: American Educational Research Association, 1974. (ERIC Document Reproduction Service No. ED 090 314)

Hater, M.A., & Kane, R.B. The cloze procedure as a measure of mathematical English. The Journal for Research in Mathematics Education, March, 1975, 6(2), 121-127.

Hayford, P.D., Schuder, R.T., & Kidder, S.J. Conventional and Rasch analyses of an unconventional test of literal comprehension. New York: New York State Department of Education, 1976.

Hittleman, D.R. Readability, readability formulas and cloze; selecting instructional materials. The Journal of Reading, November, 1978, 117-122.

Horton, R.J. The construct validity of cloze procedure: An exploratory factor analysis of cloze, paragraph reading, and structure-of-intellect tests. Reading Research Quarterly, 1974-75, 10(2), 248-251.

Jongsma, E. The cloze procedure as a teaching technique. Delaware: The International Reading Association, 1971.

Kidder, S.J. Comparing Rasch calibration of passage reading difficulty with Spache or Dale-Chall passage readability estimates. New York: New York State Education Department, 1977.

Kintsch, W., & Keenan, J. Reading rate and retention as a function of the number of propositions in the base structure of sentences. Cognitive Psychology, 1973, 5(3), 257-274.

Klare, G.R. Assessing readability. Reading Research Quarterly, 1974-1975, 10(1), 62-102.

Leithwood, K. The development of procedures for measuring competence in reading and writing. Toronto: Ministry of Education, 1 & 2, 1977.

Lewy, Arieh. Handbook of curriculum evaluation. New York: Longman Inc., 1977.

Ligon, G. A study of the relationship between affective and achievement measures. New York: American Educational Research Association, 1977. (ERIC Document Reproduction Service No. ED 137 361)

Linden Board of Education, New Jersey. Criterion reading instructional project (CRIP). Washington,D.C.: Bureau of Elementary and Secondary Education, 1975. (ERIC Document Reproduction Service No. ED 119 132)

Lorge, I. Readability formulae - An evaluation. Elementary English, 1949, 26, 86-95.

Lorge, I.D. Lorge and Flesch readability formulas: Correction. School and Society, 1948, 67, 141-142.

Lunstrum, J.P. Reading in the social studies: A preliminary analysis of recent research. Social Education, 1976, 40(1), 10-18.

MacGinitie, W.H. Contextual constraint in English prose paragraphs. Journal of Psychology, 1961, 51, 121-130.

Maryland State Department of Education, Baltimore. Division of Instruction. Standards for successful reading programs. Washington, D.C.: Office of Education, 1976. (ERIC Document Reproduction Service No. ED 119 151)

Maxwell, Martha. Readability: Have we gone too far? The Journal of Reading, May, 1978, 22, 525-530.

McLaughlin, G.H. SMOG Grading - A new readability formula. The Journal of Reading, May, 1969, 12(8), 639-646.

McNally, J. , & Murray, W. Key words to literacy. London: Schoolmaster Publishing Company, 1962.

Mehrens, W.A., & Lehmann, I.J. Measurement and evaluation in education and psychology. New York: Holt, Rinehart and Winston, 1973.

Miller, L.R. Predictive powers of multiple-choice and cloze derived readability formulas. Reading Improvement, 1975, 12(1), 52-58.

Mork, T.A., Closing the placement gap:A new tool for administrators and teachers. Educational Leadership, April, 1971, 28(7), 763-767.

Myers, Phyllis C. The cloze procedure: Latest research and uses. Paper presented at the 21st Annual Meeting of the International Reading Association , Anaheim, California, May, 197 .

Neville, M.H., & Pugh, A.K. Context in reading and listening: A comparison of children's errors in cloze tests. British Journal of Education Psychology, 1974, 44(3), 224-232.

New York State Education Department. Speed cloze exercise in a multiple-choice format. New York: New York Education Department, December, 1975.

Noble, C.E. An analysis of meaning. Psychological Review, 1952, 59, 421-430.

North York Board of Education. Alternative techniques for evaluating reading in the classroom. Toronto: North York Board of Education, 1977.

Nystrand, Martin (Ed.) Language as a way of knowing: A book of readings. Toronto: The Ontario Institute for Studies in Education, 1977.

O'Reilly, Robert P., Schuder, R.T., & Kidder, Steven J. Validation of a multiple-choice cloze test of literal comprehension. Albany, New York: New York State Education Department, February, 1976.

The Ontario Institute for Studies in Education. RHETORIC: A unified approach to English curricula. Toronto: The Ontario Institute for Studies in Education, 1970.

O'Reilly, R.P., & Streeter, R.E. Reports on the development and validation of a system for measuring literal comprehension in a multiple-choice cloze format: Preliminary factor analytic results. New York: New York State Education Department, 1976.

Otto, W., & Chester, R.D. Objective-based reading. Don Mills: Addison-Wesley Publishing Co., 1976.

Pacific Training and Technical Assistance Corporation, Berkeley, California. Product evaluation of Oakland right to read. California: Oakland Public Schools, 1973. (ERIC Document Reproduction Service No. ED 127 595)

Page, W. The post oral reading cloze test: New link between oral reading and comprehension. Journal of Reading Behaviour, 1975, 7(4), 383-389.

Panackal, A.A., & Heft, C.S. Cloze technique and multiple choice technique reliability and validity. Educational and Psychological Measurement, Winter, 1978, 38(4), 917-932.

Paradis, Edward, & Bayne, Mina. An examination of the use of cloze tasks during primary grade reading instruction. Paper presented at the 20th Annual Meeting of the International Reading Association , New York City, May 13-16, 1975.

Paton, James M. The many faces of literacy. Paper presented at the Annual Conference of the Canadian Society for the Study of Education at the University of New Brunswick, Fredericton, June 7-9, 1977.

Pikulski, J.J. Using the cloze technique. Language Arts, 1976, 53(3), 317-318, 328.

Popham, W.J. (Ed.) Evaluation in Education. California: McCutchan Publishing Corp., 1974.

Porter, D. Modified cloze procedure: A more valid reading comprehension test. English Language Teaching Journal, 1976, 30(2), 151-155.

Rankin, E.F. An evaluation of the cloze procedure as a technique for measuring reading comprehension. Dissertation Abstracts, 1958, 19(4), 733-734.

Rankin, E.F. The cloze procedure - its validity and utility. In O.S. Causey and W. Eller (Eds.), Eighth Yearbook of the National Reading Conference, 1959, National Reading Conference. Reprinted in R. Farr (Ed.), Measurement and Evaluation in Reading. New York: Harcourt, Brace and World, 1970.

Rankin, E.F. Cloze procedures- A survey of research. Yearbook of the Southwest Reading Conference, 1965,14, 133-148.

Rankin, E.F. Grade level interpretation of cloze readability scores. Paper presented at the National Reading Conference, St.Petersburg, December 3-5, 1970.

Sax, Gilbert. Principles of educational measurement and evaluation. California: Wadsworth Publishing Company, Inc., 1974.

Sigel, Irving E., & Cocking, Rodney R. Cognitive development from childhood to adolescence: A constructivist perspective. Toronto: Holt, Rinehart and Winston, 1977.

Smith, Frank. Comprehension and learning. Toronto: Holt, Rinehart and Winston, 1975.

Smith, Frank. Psycholinguistics and reading. Toronto: Holt, Rinehart and Winston, 1973.

Smith, F. Some limitations on reading and learning to read. English Quarterly, 1972,5, 45-47.

Smith, F. The role of prediction in reading. Elementary English, March, 1975, 52(3), 305-311.

Solo, Leonard. What we do because testing doesn't work. National Elementary Principal, 1977, 56(5), 63-66.

Southgate, Vera, & Roberts, G.R. Reading - which approach? London: University of London Press , 1970.

Spache, G. Contributions of allied fields to the teaching of reading. In H.M. Robinson (Ed.), Innovation and change in the reading instruction, 1968. Sixty-seventh yearbook of the N.S.S.E. Part II, 237-290.

Spreen, O., & Schulz, R. Parameters of abstraction, meaningfulness, and pronounciability for 329 nouns. Journal of Verbal Learning and Verbal Behaviour, 1966, 5, 459-468.

Standards of excellence for reading programs in Iowa schools, Des Moines: Iowa State Department of Public Instruction, 1976.

Stevens, John. Is there a literate in the house? Forum, December, 1977, 237-242.

Taylor, Insup. Introduction to psycholinguistics. Montreal, Toronto: Holt, Rinehart and Winston, 1976.

Taylor, Wilson L. Cloze procedure: A new tool for measuring readability. Journalism Quarterly, Fall, 1953, 30, 415-433.

Taylor, Wilson L. Recent developments in the use of the cloze procedure. Journalism Quarterly, 1956, 33, 42-48.

Taylor, Wilson L. Cloze readability scores as indices of individual differences in comprehension and aptitude. Journal of Applied Psychology, 1957, 41, 19-26.

Thelen, J.N. Using the cloze test with science textbooks. Science and children, 1974, 12(3), 26-27.

Theobald, John T., & Alexander, J. Estill. An auditory cloze procedure for assessing the difficulty level of teacher instructional talk in the intermediate grades. The Elementary School Journal, May, 1977, 389-394.

Tighe, Donald J. Ways of communicating, An anthology. Massachusetts: Winthrop Publisher, Inc., 1975.

Tovey, D.R. The psycholinguistic guessing game. Language Arts, March, 1976 , 53(3), 319-325.

Tracy, Patricia, & Campbell, Ellen. A look at high school reading assessment from a Canadian viewpoint. Toronto: The Ontario Institute for Studies in Education, 1973.

Trent Valley Centre. The eastern Ontario language arts measurement project - writing (K-10). Toronto: The Ontario Institute for Studies in Education, 1976.

Tuinman, J.J., Blanton, W.E., & Gray, G. A note on cloze as a measure of comprehension. Journal of Psychology, 1975, 90, 159-162.

Walter, Richard B. Historical overview of the cloze procedure. New Jersey: Kean College, 1974.

Weintraub, R. The cloze procedure. The Reading Teacher, 1968, 21, 567-571.

Williamson, H.J., & Ball, I.L. Readability levels of children's literature, Melbourne: Educational Resources, 1973.

Winkeljohann, Rosemary. A selective bibliography of ERIC abstracts for the teacher of reading. Illinois: ERIC Clearinghouse on Reading and Communication Skills, 1976. (ERIC - ED 127 603)